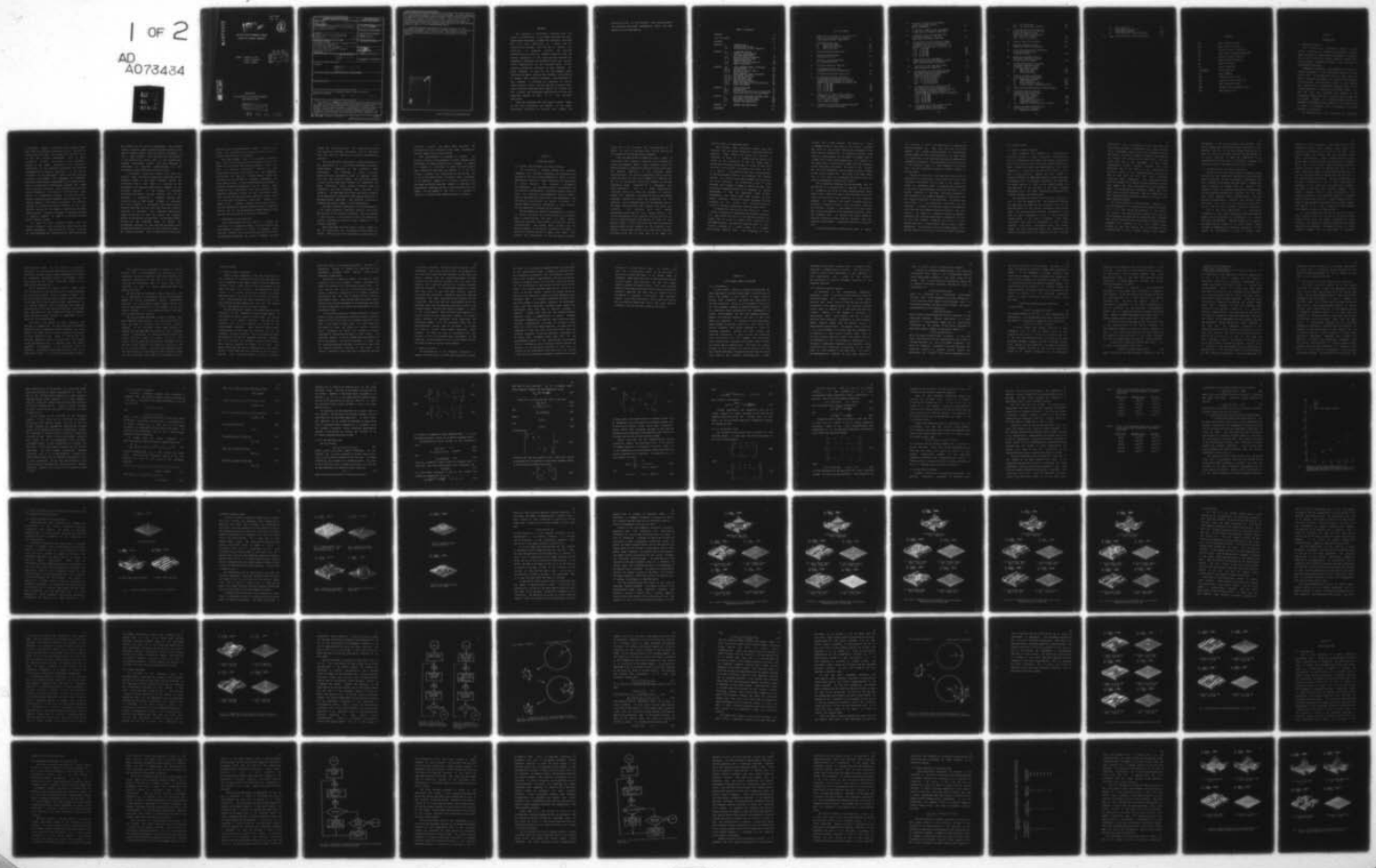
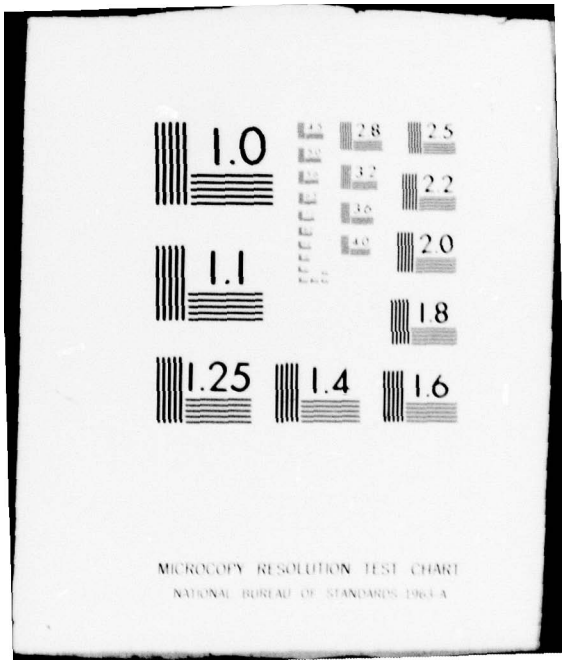
AD-A073 434 UTAH UNIV SALT LAKE CITY F/G 3/1
HIGH RESOLUTION ASTRONOMICAL IMAGING THROUGH THE TURBULENT ATMO--ETC(U)
JUL 79 R L FROST, C K RUSHFORTH, B S BAXTER AFOSR-77-3212
UNCLASSIFIED UTEC-79-081 NL

1 OF 2
AD A073434




1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD A073434

UTEC 79-081

Copy 9

LEVEL III

HIGH RESOLUTION ASTRONOMICAL IMAGING THROUGH THE TURBULENT ATMOSPHERE

Authors: Richard L. Frost
Craig K. Rushforth
Brent S. Baxter

DDC RECEIVED AUG 31 1979 C

DDC FILE COPY

Sponsored by

Air Force Office of Scientific Research
Grant AFOSR 77-3212

This document has been approved for public release and sale; its distribution is unlimited.

79 08 30 020

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UTEC-79-081
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): "High Resolution Astronomical Imaging Through The Turbulent Atmosphere.
5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Richard Lee Frost, Craig K. Rushforth, Brent Baxter
8. CONTRACT OR GRANT NUMBER(s): Grant AFOSR-77-3212
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Utah, Departments of Electrical Eng. & Computer Science, Salt Lake City, Utah 84112
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Bolling Air Force Base, Washington, D.C.
12. REPORT DATE: July 1979
13. NUMBER OF PAGES: 120
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 132 p
15. SECURITY CLASS. (of this report)
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): unlimited

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Atmospheric turbulence; astronomical imaging; image restoration, speckle interferometry

79 08 30 020

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research is principally concerned with the digital reconstruction of star images observed with large ground-based telescopes, although the techniques developed here will have application to a broad class of reconstruction problems. Since the work of Labeyrie, the difficulty in producing accurate and detailed reconstructions of stars has stemmed primarily from the extreme degradation of the phase spectrum caused by the atmospheric turbulence, and secondarily from the low-pass filter characteristic of the telescope itself.

CONT

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

362700

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

In this research, we describe solutions to both problems, Our phase estimator is based on the Knox-Thompson phase difference estimator, which we have extended and modified to produce more accurate estimates. The performance of this estimator is evaluated by simulation at various signal-to-noise ratios. We also describe a new non-linear super-resolution algorithm which appears to exhibit the best accuracy and convergence characteristics of any such algorithm proposed to date. It is also evaluated empirically.

These two techniques were then used to restore images of the stars Betelgeuse and Capella. In the latter restoration resolution is realized well beyond the diffraction-limit of the telescope. Both reconstructions are consistent with known astrophysical facts, and both appear to be of high quality.

Accession For
NTIS G.A&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

ABSTRACT

This research is principally concerned with the digital reconstruction of star images observed with large ground-based telescopes, although the techniques developed here will have application to a broad class of reconstruction problems. Since the work of Labeyrie, the difficulty in producing accurate and detailed reconstructions of stars has stemmed primarily from the extreme degradation of the phase spectrum caused by the atmospheric turbulence, and secondarily from the low-pass filter characteristic of the telescope itself. In this research, we describe solutions to both problems. Our phase estimator is based on the Knox-Thompson phase difference estimator, which we have extended and modified to produce more accurate estimates. The performance of this estimator is evaluated by simulation at various signal-to-noise ratios. We also describe a new non-linear super-resolution algorithm which appears to exhibit the best accuracy and convergence characteristics of any such algorithm proposed to date. It is also evaluated empirically.

These two techniques were then used to restore images of the stars Betelgeuse and Capella. In the latter restoration resolution is realized well beyond the

diffraction-limit of the telescope. Both reconstructions are consistent with known astrophysical facts, and both appear to be of high quality.

TABLE OF CONTENTS

ABSTRACT			iv
LIST OF FIGURES			vii
GLOSSARY			xi
CHAPTER	I	INTRODUCTION	1
	I.1	Problem Description	1
	I.2	Contribution of this Research	4
CHAPTER	II	LITERATURE REVIEW	7
	II.1	Speckle Interferometry and some extensions	7
	II.2	Explicit Phase Estimation	12
	II.3	Estimating Phase from Phase Differences	14
	II.4	Implicit Phase Estimation	18
	II.5	Super-resolution	20
CHAPTER	III	LEAST-SQUARES PHASE ESTIMATION	23
	III.1	Introduction	23
	III.2	Review of Speckle Imaging	24
	III.3	Derivation of Least-Squares Phase Integrator	29
	III.3.1	The Normal equations	31
	III.3.2	The Aperiodic case	33
	III.3.3	The Periodic case	37
	III.4	Algorithm performance	39
	III.5	Phase Estimation on Simulated Data	44
	III.6	Phase Refinement	59
CHAPTER	IV	SUPER-RESOLUTION	71
	IV.1	Introduction	71
	IV.2	Background and Analysis of Algorithm	72
	IV.3	Super-resolution of Simulated Data	81
CHAPTER	V	RESTORING TURBULENCE-BLURRED STARS	98
	V.1	Description of Data Base	98
	V.2	Betelgeuse reconstruction	99
	V.3	Capella reconstruction	105
CHAPTER	VI	SUMMARY AND CONCLUSIONS	112
REFERENCES			115

LIST OF FIGURES

1	Comparison of growth of reconstruction error vs. array size for various least-squares phase integrators	43
2	Simulated binary star	
	a) Intensity object	45
	b) Magnitude spectrum	45
	c) Phase spectrum	45
3	Uncorrelated Gaussian noise	47
4	Gaussian spatial phase correlation filter	47
5	Spatially correlated phase of instantaneous OTF	47
6	Telescope aperture function	47
7	A simulated PSF for the atmosphere-telescope combination	48
8	A simulated blurred binary star	48
9	Reconstructed phases and their corresponding reconstructed objects using both line-integral and least-squares phase estimates at various SNR	
	a-e) 5 dB SNR	51
	f-j) 10 dB SNR	52
	k-o) 20 dB SNR	53
	p-t) 30 dB SNR	54
	u-y) 40 dB SNR	55
10	Comparison of least-squares phase estimates and object reconstructions for two different spatial windows	
	a-b) Fourier window	60
	c-d) Hamming window	60
11	Flowchart of Gerchberg-Saxton algorithm used in electron microscopy	62

12	Flowchart of modified Gerchberg-Saxton algorithm used for phase refinement	62
13	A geometric proof of the convergence of the phase refinement algorithm	63
14	A geometric proof of the possible non-convergence of a similar magnitude refinement algorithm	67
15	Illustration of the effects of phase refinement on the least-squares phase estimates and their corresponding object reconstructions at various SNR	
	a-b) 5 dB SNR	69
	c-d) 10 dB SNR	69
	e-f) 20 dB SNR	69
	g-h) 30 dB SNR	70
	i-j) 40 dB SNR	70
16	Flow chart of the Gerchberg super-resolution algorithm including the 'noise-limiting' procedure	75
17	Flow chart of the new non-linear super-resolution algorithm	78
18	Illustrations of the effects of the new super-resolution algorithm on noise free data at various initial cutoff frequencies	
	a-d) 20 cycles/picture	84
	e-h) 15 cycles/picture	85
	i-l) 10 cycles/picture	86
	m-p) 7 cycles/picture	87
19	Illustrations of the effects of the new super-resolution algorithm on the original spectrum estimates and the corresponding reconstructions at various SNR and initial cutoff frequencies	
	a-e) 5 dB SNR	88
	f-j) 10 dB SNR	89
	k-o) 20 dB SNR	90
	p-t) 30 dB SNR	91
	u-y) 40 dB SNR	92
20	Illustrations of the effects of the Gerchberg super-resolution algorithm on the 5 dB SNR data	

	a-c) 150 iterations	94
	d-f) 29 iterations (optimal)	94
21	Illustrations of the effects of overestimating the size of the region of support on the super-resolved reconstructions	
	a-c) Area of 24 pixels	96
	d-f) Area of 32 pixels	96
22	Speckle image of Betelgeuse	100
23	Labeyrie estimate of the autocorrelation of Betelgeuse	100
24	Least-squares phase of Betelgeuse obtained without using a spatial window	100
25	Magnified reconstruction of Betelgeuse obtained without using a spatial window	100
26	Estimated Betelgeuse spectrum and object obtained using a Hamming spatial window	
	a) magnitude spectrum	101
	b) phase spectrum	101
	c) intensity object	101
27	Effects of super-resolution reconstruction of Betelgeuse	
	a) magnitude spectrum	103
	b) phase spectrum	103
	c) line drawing of object	103
	d) object radial intensity profile	103
28	Super-resolved Betelgeuse intensity object	104
29	Estimated Capella spectrum and object obtained using a Hamming spatial window	
	a) magnitude spectrum	109
	b) phase spectrum	109
	c) antidiagonal profile of magnitude spectrum	109
	d) Labeyrie autocorrelation estimate	109
30	Effects of super-resolution on reconstruction of Capella	
	a) magnitude spectrum	110

	b) phase spectrum	110
	c) antidiagonal profile of magnitude spectrum	110
	d) line drawing of intensity object	110
31	Super-resolved Capella intensity object	111

GLOSSARY

PSF	point spread function
PSD	power spectral density
DFT	discrete Fourier transform
FFT	fast Fourier transform algorithm
OTF	optical transfer function
ACF	autocorrelation function
FA	Fourier analysis
CR	cyclic reduction
ISOPLANATIC	Shift-invariant
K-T	Knox-Thompson
SNR	signal-to-noise ratio
SOR	successive over-relaxation
SSE	sum of squared errors
WSSE	weighted sum of squared errors
MSE	mean square error

CHAPTER I

INTRODUCTION

I.1 Problem Description

The restoration of astronomical images to permit greater resolution is a problem of long-standing interest. Image resolution is limited both by atmospheric turbulence and by the diffraction effects caused by imaging with a finite aperture, although the former is more serious for ground-based telescopes.

Atmospheric turbulence can be thought of as causing a random perturbation of the refractive index of air in the optical path. This perturbation varies in both time and space, and causes the phase of an incident light wavefront to be distorted. These distorted waves are not brought into sharp focus by a conventional imaging system, but are instead spread over a large area in the image plane. For the 200 inch Hale telescope, the diameter of the diffraction-limited Airy disk in visible light is about .05 arc-seconds. However, for observations made through the atmosphere the minimum signal diameter in the image plane is about 2 arc-seconds, which corresponds to a 40-fold decrease in resolution.

For many years this lost resolution was considered

irrecoverable. However, in the early 1970s Labeyrie showed that diffraction-limited information could be obtained by ground-based telescopes. His approach was to take a series of very short exposure photographs, each of which essentially 'freezes' the motion of the atmosphere. Because of the random constructive and destructive interference of the distorted light waves, the resulting photographs have a speckled appearance. If the Fourier transform of each photograph is taken, the phase spectrum is seen to vary randomly from photograph to photograph. When these short-exposure photographs are averaged directly (which is equivalent to long-exposure imaging), the random phase spectrum variations cause cancellation of the high spatial frequencies. The result is a low-resolution image typical of conventional astronomy. To avoid this result, Labeyrie ignored the phase component and averaged only the squared modulus of the Fourier transforms. The resulting estimate is a diffraction-limited estimate of the power spectral density of the image, or equivalently the image autocorrelation. Because no estimate of the phase spectrum is obtained, however, an image of an arbitrarily shaped object cannot be formed.

To supply the phase spectrum estimate, Knox and Thompson proposed a new technique for processing the same speckle photographs. They multiply the Fourier transform of each photograph by a shifted copy of itself and average

this product over the series of photographs. This averaged product estimates phase differences between points in the spatial frequency domain, which can then be integrated to form a phase estimate. Knox indicates that the estimator is very sensitive to sensor noise in the observations. Although encouraging results have been obtained in simulations, no successful reconstructions of real astronomical objects have been obtained to date.

As mentioned, diffraction effects also limit the possible image resolution by imposing an absolute spatial frequency cutoff on the signal spectrum, called the diffraction limit. This limit is a function of the wavelength of the observed light and the diameter of the telescope aperture. In the absence of telescope aberrations the optical transfer function (OTF) of the telescope is essentially a triangular function that low-pass filters an input signal. A number of techniques have been proposed to extrapolate the measured spatial frequency spectrum, the most successful being the iterative algorithm of Gerchberg. This algorithm alternatively imposes constraints in both the spatial and frequency domains to perform the extrapolation. As Gerchberg himself pointed out, the algorithm tends to corrupt the extrapolated spectrum at each iteration by adding to the extrapolated spectrum a portion of the distortion energy in the measured spectrum. Thus, extrapolations of noisy data

obtained by this algorithm may be suspect. Despite this caveat, the algorithm has recently been used to aid in a digital reconstruction of Betelgeuse.

These two problems, that is, accurately estimating phase spectra and extrapolating low-frequency spectra, are common to a great many deconvolution problems. There is to date no entirely satisfactory solution to this deconvolution problem in the general case. The fundamental difficulties are those of inversion theory; namely, the truncation of infinite impulse responses, smoothing, and noise amplification. In the blind deconvolution problem, a spectral estimation problem must also be solved. Phase estimation has traditionally been more difficult than magnitude estimation, so much so that researchers often ignore it when dealing with speech and music signals, where it is often not perceptually significant. However, for images, such a cavalier approach leads to poor results. Thus the astronomical imaging problem presents an important special case of the deconvolution problem because of the poor SNR and the severity of the phase distortion.

I.2 Contribution of this Research

This research describes our efforts to improve the resolution of astronomical images, first by modifying the Knox-Thompson estimator to provide an adequate phase spectrum estimate out to the telescope diffraction-limit, and second by estimating the spatial frequency spectrum

beyond the diffraction-limit. The techniques described here have been strikingly successful, and we anticipate that they will be applied to a variety of deconvolution problems.

Our most important contribution to phase estimation is the application of a fast direct algorithm to perform a least-squares integration of the estimated phase differences. The algorithm is stable, accurate, memory-efficient, and about an order of magnitude faster than the best iterative schemes. The resulting estimates are more accurate than those obtained by integrating the phase differences along several different paths of integration and then averaging these discrete line integrals. We show empirically that even better estimates can be obtained by refining the least-squares phase with a Gerchberg-Saxton algorithm. The practical necessity of windowing the speckle images is also demonstrated.

The other principal contribution is the development of a new super-resolution algorithm, based on a non-linear modification of the Gerchberg algorithm. This new approach is more accurate in the presence of noise and demonstrates superior convergence properties when compared to the original algorithm.

These techniques have been used to restore images of the red super-giant star Betelgeuse and the binary star Capella. These are the first successful reconstructions to

explicitly estimate the object phase spectrum. The reconstructions are consistent with known astrophysical data and appear to be of high quality.

This dissertation is arranged as follows. The previous research on the problems of speckle imaging, phase estimation, and super-resolution are reviewed in chapter II. Chapter III describes in detail the least-squares phase integration algorithm, and compares it to the line integral estimator via simulations for various SNR. The use of various modifications, including phase refinement, is also explored. Chapter IV is a similar treatment of the new super-resolution algorithm. Chapter V describes the experimental technique and the results obtained using real images. A brief summary of the results is reviewed in chapter VI, including some remaining areas of research.

CHAPTER II

LITERATURE REVIEW

II.1 Speckle Interferometry and Some Extensions

The difficulty in imaging through the turbulent atmosphere has long been recognized, and astronomers have resorted to the use of interferometers to obtain accurate quantitative data on the angular dimensions and separations of stellar objects. The classical astronomical interferometers, the Michelson stellar interferometer and the intensity interferometer, invented by Hanbury Brown and Twiss, are inherently one-dimensional devices, however. They are therefore most useful when a one-dimensional scan sufficiently characterizes the object of interest. For extended objects pictorial information is often desired, and one-dimensional techniques do not suffice.

The development of speckle interferometry by Labeyrie [1] and his successful use of the procedure [2] to measure object autocorrelations and hence to infer angular dimensions represent a step towards achieving full aperture reconstructions. The method, which averages only the squared modulus of the Fourier transform of the image, is successful precisely because it ignores the phase. This prevents the cancellation of the high spatial frequency

energy, but it also precludes the reconstruction of an image. As such it represents only a partial solution to the problem of diffraction-limited imaging.

Since the publication of Labeyrie's work a number of researchers have proposed extensions to the technique to permit true imaging. Most of these proposals avoid the problem of phase estimation by working directly with either the speckles themselves or the estimated object autocorrelation. The earliest suggestions (those of Bates, Gough and Napier [3] and Liu and Lohmann [4]) required the use of an unresolvable point star somewhere in the same isoplanatic patch. (Isoplanicity is synonymous with shift-invariance). This is a rather unsatisfactory criterion, since it is often not possible in practice. Weigelt [5] simulated a variation of the Liu and Lohmann technique, wherein he varied the brightness of the reference star to determine which of the apparent stars in the image autocorrelation were genuine and which were artifacts. The artifacts were then removed photographically during the final copying process. The relative brightness of the reference star was varied by the use of a neutral density filter placed in front of each speckle image. In fact, the filter was used to darken the speckles which are not produced by the reference star. Weigelt reported good results in his simulation, even when the object field was so large that it no longer lay

entirely within an isoplanatic patch.

More recently, some interesting results have been obtained from real speckle images by an entirely different approach. Lynds, Worden and Harvey [6] describe a technique in which it is assumed that the speckles are formed by the convolution of the true image with a point-spread function consisting of an array of Dirac delta functions, each of which is located at the center of mass of a speckle. The problem is then one of deconvolution. An editing procedure selects the brightest and clearest speckles so as to most accurately estimate the location of the delta functions. The deconvolved result is then averaged over a number of images to improve the SNR. They carried out a demonstration of this procedure on speckle photographs of the red supergiant Alpha Orionis (Betelgeuse), and indeed found some evidence of fine surface structure. There were internal evidences that the technique did not achieve full diffraction-limited resolution, however, despite the increased image detail.

To improve the resolution of the Lynds et al. result, a further refinement was implemented by McDonnell and Bates [7]. They accept the result as a preliminary estimate, but assume that an array of simple delta functions is an oversimplification, and that the resulting image is still blurred, although to a lesser extent, by a residual point-spread function (PSF). The estimation of this

residual PSF is rather involved. The Lynds et al. result is assumed to have as the zeros of its Fourier transform the union of the zeros of the Fourier transforms of the object and the residual PSF. To separately identify these zeros an intensity model for the object is assumed, that of a simple first order limb-darkened star. The identity of the image zeros is determined by comparison with reasonable values predicted by the model. The zeros of the residual PSF are thus identified, and the image is Weiner filtered by the inverse residual PSF. The result is an image whose size is very near the telescope diffraction limit, and certain astrophysical features are visible.

This restoration technique has the advantage that no observation of a reference star is required. The Lynds et al. assumption only holds for small bounded objects whose angular size is about the same as the telescope diffraction limit, which restricts the usefulness of the technique somewhat. The McDonnell and Bates refinement would be difficult to extend to objects with non-simple intensity distributions, because a more complex set of zeros might defy an easy segregation of PSF and object transform zeros into separate sets. In addition, the need to assume an intensity distribution model begs the question since the intensity distribution is precisely what we are trying to find.

A related processing technique which seeks to remove

the necessity of point star observation is suggested by Welter and Worden [8]. They show that the difference between the average autocorrelation of instantaneous speckle images and the crosscorrelation of the same images is proportional to the image autocorrelation undegraded by atmospheric turbulence. The approximation is most accurate if the observed object is small compared to the seeing spot. How this technique may apply to speckle imaging is not yet known.

Sherman [9] describes a technique whose application is unhindered by restrictions on the location of a point star reference or the object size, so long as the image lies within an isoplanatic patch. He proposes the calculation of the entire four-dimensional covariance function of the Fourier transform of the image. The object transform is then estimated as the complex eigenmatrix associated with the largest eigenvalue of the covariance. He obtains very good results in a one-dimensional simulation, and the technique is claimed to be much more immune to measurement noise than the Knox-Thompson method discussed in the next section. The amount of computation needed to implement completely this technique would make it impractical for all but the most powerful computing facilities. In [10] Sherman describes a modification which greatly reduces computation. At the present time, his method appears to be the most practical alternative to explicit phase estimation

for a general image.

II.2 Explicit Phase Estimation

Phase estimation has remained an extraordinarily difficult problem in image restoration. Recent work by Cole [11] demonstrates the rather poor quality restorations which result when real world blurs are assumed to have zero phase. Cannon [12] showed some improvement by the simple expedient of considering the phase to be either 0 or pi, that is, phase reversals. However, the first attempt to actually estimate the image phase appears to have been done by McGlamery [13]. He averaged separately the magnitude and phase spectra of each of the speckle images. He concluded that as the variance of the phase estimate approaches pi/3, the phase information provided by this estimator becomes vanishingly small.

The technique of central concern to this research is that of Knox and Thompson [14]. They proposed estimating the image phase by computing the one-shift term of the autocorrelation of the Fourier transform of each speckle image. This term preserves the phase information in the form of phase differences in the x- and y-directions between points in the Fourier transform. These phase differences must then be integrated to form a phase estimate. In his Ph.D. dissertation Knox [15] further showed by both analytical means and simulations the relative insensitivity of the technique to fixed telescope

aberrations. This is fortunate, since it may permit the use of this technique in conjunction with large multiple mirror telescopes, which often have substantial optical aberrations. A similar extension of speckle interferometry has already been worked out by Roddier [16]. Knox also found that, due to the narrowness of the random instantaneous optical transfer function (OTF), the normalized SNR of the terms in the autocorrelation decreases rapidly as the shifted distance increases, so that at best only the first few terms contribute significant phase information. Knox suggested that the phase be estimated by averaging the phase estimates derived from summing the phase differences over two independent paths. His one-dimensional simulations showed that very good results could be obtained. They also showed that the final reconstructions were quite sensitive to errors in the phase spectrum estimate.

At Itek Corporation a two-dimensional simulation of the Knox-Thompson (K-T) technique was performed by Nisenson et al. [17]. They confirm the potential of the method, and indicate that improved phase estimates can be derived from averaging a multiplicity of integrated paths, although they do not state which paths were used, nor how many were averaged. R.V. Stacknik et al. [18], another group working at Itek then applied the K-T technique to the reconstruction of solar surface features from speckle

photographs, a test case providing a very high SNR. Their results showed near diffraction-limited resolution on data obtained from the 24 inch solar vacuum telescope at Kitt Peak, AZ. Both of these papers mention the development of an optical processor to speed the computation of the autocorrelation of the Fourier transforms, although the processing actually implemented was entirely digital.

II.3 Estimating Phase from Phase Differences

We have already noted that the phase estimation scheme proposed by Knox makes only limited use of the available phase information, and that a substantial improvement has been reported by averaging over many paths. Attempting a direct extension to average over all possible line integrals leads to an unwieldy combinatorial problem whose solution would be prohibitively expensive for the large data arrays (256x256) being considered here. This same problem of optimum phase estimation from measured phase differences also arises in the design of active optical devices, and is in fact a problem in numerical integration.

One of the first applications of least-squares estimation to this phase integration problem was made by Rimmer [19] in evaluating the performance of a lateral shearing interferometer. Shortly thereafter, his approach was applied to the design of an active optical system intended to compensate for atmospheric turbulence. In this system, an interferometer is used to measure phase

differences across an aperture. These differences are then integrated to form a phase estimate, whose effects are removed by a deformable lens or mirror in real time. Fried [20] and Hudgin [21] derived the same equations again almost simultaneously. Hudgin assumed that the solution was periodic, as when using the Discrete Fourier transform (DFT), while Fried did not. Both researchers derived an iterative solution scheme, and demonstrated that for the small array sizes common in active optics the rms error in the final phase estimate was smaller than the rms error in the measured phase differences. They also showed that the rms phase reconstruction error is only logarithmically dependent on the array size, indicating that it might be applicable to the large arrays of speckle imaging. Fried further reported that to assure convergence of his iterative solutions, he increased the number of iterations almost linearly with array size. Based on Hudgin's work, a group at Itek headed by Hardy [22] built an analog computer and applied it to a small (about 5x4 array) prototype active optical system, with good results.

In a recent paper Hunt [23] derives the matrix form of these equations and points out that the iterative solution scheme of both Fried and Hudgin corresponds to Jacobi iteration, the numerical iteration scheme with the slowest known rate of convergence. He suggests the use of faster iterative schemes, such as Gauss-Seidel or successive

over-relaxation (SOR). He also derives an analytical expression for the rms phase reconstruction error based on a periodic structure which slightly underbounds the results of simulations obtained by both Fried and Hudgin. Hunt's result shows the same nearly logarithmic growth of rms phase error with array size.

Hunt's suggestion of a faster iterative scheme leads one to speculate on the possible use of a direct algorithm for solving the problem. As will be shown in chapter III, the least-squares equations have a highly regular structure of the same form as the 5-point discrete approximation to the Poisson equation. This is fortunate, since numerical analysts have in the last decade devised some surprisingly fast algorithms for the solution of Poisson's equation. There are three distinct classes of algorithms, all of which will be discussed below.

The first is the method of Cyclic Reduction (CR) devised by Buneman [24]. He noted that the even-numbered equations could be combined to form an equation for every fourth line, which could in turn be reduced to an equation for every eighth line, etc. There is finally only one equation remaining, which is easily solved. The solution on the odd-numbered lines is obtained from the reduced equations corresponding to the successive levels of reduction. The method is particularly easy to code, and is numerically stable. It is not the fastest method, however.

The second class of algorithms are based on Fourier analysis (FA) or transformation of the data so as to diagonalize the matrix equations. When these methods have dimensions appropriate to the use of a Fast Fourier Transform (FFT) algorithm, they are faster than cyclic reduction, as well as being memory-efficient and numerically stable.

The third type of algorithm is that of Giwa [25], and is an adaptation of an analytical variation-of-parameter method for solving second-order partial differential equations. It is very general, highly stable, and does not restrict the dimensionality of the problem, as do the Fourier methods. It is less memory-efficient and slower than the other methods, however.

The fastest known algorithm is that of Hockney [26], which combines both FA and CR. It is the most difficult to code, and must be carefully used to avoid overflow problems. In his excellent review paper, Hockney [27] compares his own algorithm with CR and the best iterative schemes, including SOR. He indicates that his FACR is the fastest, that FA is about 50% slower, and CR is 100% slower. He also shows that to guarantee a maximum error of one part in 10E6, about 432 iterations are required of SOR on an array of 128x128. Since the direct methods achieve the same accuracy in a time equal to that required by about six SOR iterations, there seems to be no reason to use an

iterative method.

II.4 Implicit Phase Estimation

Implicit phase estimation is the name I have given to the technique of estimating the phase spectrum based on a priori knowledge of the signal properties and the magnitude spectrum alone, rather than on direct measurements of the phase spectrum or its derivative. The approach has appeal because it is usually much easier to measure accurately magnitude spectra, as in the case of electron microscopy, x-ray crystallography, and image restoration [30]. It has been explored by many authors over the past two decades [31-36], and remains an area of active interest.

For one-dimensional signals, the problem may be easily understood in the framework of linear system theory. If the duration of such a signal is finite, then its Fourier transform is an analytic function. Thus, the real and imaginary parts of the transform are related by the Cauchy-Riemann equations. Other analytic spectra may be formed however by reflecting any transform zero to its conjugate reciprocal location. This will modify the phase spectrum while leaving the magnitude spectrum unchanged. Wolf [33] derived the allowable forms of the resulting phase spectra assuming the object to be real, thus demonstrating that the object reconstruction is not unique. However, in two dimensions the situation is not as easy to describe. This is because analysis in terms of transform

poles and zeroes is not generally possible. The lack of a fundamental theorem of algebra for functions of two independent variables makes spectral factorization techniques inapplicable.

Nevertheless a number of schemes for implicit phase estimation have been proposed. Kohler and Mandel [32] compare two methods on one dimensional spectral data. The first involves the use of a reference point source, the second a numerical integration of a Hilbert transform relationship when the zero locations in the z plane are approximately known. They indicate that the second method requires very accurate knowledge of the magnitude spectrum, and their method does not seem to have been applied to real astronomical data.

In electron microscopy a slightly different situation obtains in that the magnitude spectrum is known in both the image and diffraction planes of an imaging system while the phase is not known in either plane. It is well known that the signals in these two planes are Fourier transform pairs. Gerchberg and Saxton [37] propose an iterative algorithm for determining the phase in both planes. The algorithm assumes an initial random phase distribution in one of the domains. The Fourier transform is taken, and the resulting magnitude spectrum is corrected to agree with the measured magnitude spectrum in that domain. The inverse transform is then taken and its magnitude spectrum

is similarly corrected. The entire process iterates until convergence, which may be quite slow. The uniqueness of the resulting solution has been examined in [38], where the authors conclude that the reconstructed phase is unique provided that the magnitude spectrum does not have even symmetry and is non-zero, and that the resulting frequency spectrum is analytic. If the magnitude spectrum is even, there may exist at most one other analytic function satisfying the constraints. Methods to speed the algorithm convergence have been described by both Gassman [39] and Feinup [40]. The latter has also experimented with this algorithm when only the magnitude spectrum is known in the frequency domain, and only the region of support is known in the signal domain. While solution non-uniqueness is a serious problem for one dimensional signals, he obtains unique solutions on simulations using complicated two dimensional images. It is not presently known why the two-dimensional results seem to be unique. It is known [41] that an initial (even highly inaccurate) phase estimate can often be used to distinguish the true solution, as well as to greatly speed convergence of the algorithm. Feinup proposes the use of this method to solve the phase problem of speckle interferometry.

II.5 Super-resolution

The extrapolation of the frequency spectrum of a lowpass filtered signal is called super-resolution, because

its effect is to increase the apparent detail or resolution of the reconstructed signal. A number of techniques have been proposed to extrapolate the low-frequency spectrum. Harris [42] proposed approximating the true spectrum by forming a weighted sum of sinc functions, where the weights are the coefficients of the discrete Fourier series of the signal. The extended spectrum appears to be sensitive to noise in the data, however. Barnes [43] expands the given portion of the spectrum in terms of prolate spheroidal wave functions [44]. It is shown by Rushforth and Harris [45] that this method is also highly sensitive to measurement error, and that achieving resolution much beyond the diffraction limit with data of realistic accuracy is doubtful. Frieden [46] describes the 'maximum entropy' approach, which is probabilistic in nature. Although substantial computation is required, good results have been obtained. Gerchberg [47] describes an iterative scheme which he calls error-energy reduction. He shows that the algorithm is linear and that given noise-free data the solution converges to the true spectrum. Convergence can be quite slow. Gerchberg notes that error in the extrapolated portion of the spectrum due to measurement noise is increased at every iteration, thereby limiting the accuracy of reconstruction possible on real data. An analysis of the algorithm by Youla [48] casts the problem in the form of alternating orthogonal projections on linear

subspaces of a parent Hilbert space. His analysis also shows the ill-conditioned nature of the algorithm. In particular he notes that there is an optimal number of iterations after which the error in the continued spectrum increases rather than decreases, and that this number cannot in general be known a priori. Still, this technique seems to be the most successful to date, and an essentially identical algorithm has been used by Papoulis [49] to extrapolate a truncated band-limited time waveform. However, a new algorithm based on Youla's work has recently been described by Cadzow [50], and it may prove superior although it has not yet been tested on noisy data.

CHAPTER III

LEAST SQUARES PHASE ESTIMATION

III.1 Introduction

In this chapter I describe a method of estimating the phase spectrum of an object whose image is both blurred and noisy. The method is very similar to power spectral density (PSD) estimation, and should have application in the same situations. The basis of the method is the use of the Knox-Thompson estimate of phase differences. Section III.2 contains a brief explanation of the fundamentals of speckle interferometry and the Knox-Thompson estimator. However, for detailed analysis, the work of Knox [15] should be consulted. In section III.3 the least-square phase estimate is derived, as well as a fast direct Fast Fourier Transform (FFT) based algorithm to solve the resulting equations. Section III.4 will discuss the algorithm performance in terms of speed and integration error, and section III.5 will compare the least-square phase estimate with a two-path line integral phase estimate for various SNR on simulated data. The effects of spatial windows on the estimates will also be noted, as will the use of phase difference information derived from the second shift term of the transform autocorrelation. Further

refinement of the phase estimates with a Gerchberg-Saxton algorithm is demonstrated in section 6. This section also contains a proof of the convergence of the algorithm in this application, and a proof of the possible non-convergence of a similar refinement operation on the magnitude spectrum.

III.2 Review of Speckle Imaging

As mentioned in the introduction, atmospheric turbulence causes a variation in the refractive index of the optical path. This is modeled by a point spread function (PSF) which varies in both the time and spatial coordinates of the imaging system. The resulting blurring process must be modeled as a complicated superposition integral. However, if the imaging is restricted to sufficiently small solid angles, the PSF is nearly spatially shift-invariant. Furthermore, for short-exposure images, Labeyrie showed that the physical process is approximately convolutional. The exposure times, typically 5 to 10 milliseconds, are so short that the atmosphere can be regarded as a fixed optical medium, whose index of refraction remains spatially random. The primary effect of the atmosphere on each short-exposure photograph is to introduce a random phase component to the spectrum, causing constructive and destructive interference of the light waves to occur randomly in the image plane. As a result, the photographs are composed of many small patches of

light, or speckles, hence the name speckle imaging.

Consider an incoherent imaging system with intensity distribution $i(\bar{x})$, where $\bar{x}$ is a 2-dimensional position vector describing a location in the focal plane of the system. Over a shift-invariant (isoplanatic) region $i(\bar{x})$ is related to the object intensity distribution $i_o(\bar{x})$ by convolution, as

$$i(\bar{x})=\int i_o(\bar{\alpha})s(\bar{x}+f\alpha/z)\,d\alpha, \qquad (1)$$

where $s(\alpha)$ represents the instantaneous point-spread function (PSF) of the atmosphere-telescope combination, f is the effective focal length of the telescope, and z is the distance between the object and telescope. The Fourier transformation of this equation is

$$I(\bar{u})=I_o^*(f\bar{u}/z)S(\bar{u}), \qquad (2)$$

where capital letters denote the Fourier transform of the lower-case quantity. Because convolution is a linear operation, a long-exposure image can be equivalently represented by the average of the Fourier transforms of many short-exposure images in the Fourier domain, as

$$\langle I(\bar{u})\rangle=I_o^*(f\bar{u}/z)\langle S(\bar{u})\rangle, \qquad (3)$$

where the angle brackets denote averaging. It has been shown both theoretically and experimentally that $\langle I(\bar{u})\rangle$ is a narrow function, i.e. has a low cutoff frequency. The heart of Labeyrie's technique is the fact that under the conditions required to produce speckle images the mean-square OTF ($\langle |S(\bar{u})|^2 \rangle$) has non-zero magnitude out to

the diffraction-limit of the telescope. The phase of the image transforms has suffered a random phase perturbation, however. If one attempts to average directly many short-exposure images or their transforms the random phase, which is most pronounced at the higher frequencies, causes a cancellation of the high spatial frequency energy. This cancellation can be avoided if one averages instead the squared modulus of the transforms, as first shown in this context by Labeyrie [1]. This averaging essentially ignores the phase contribution, whatever it may be. Thus, one forms

$$\langle I(\bar{u})I^{*}(\bar{u})\rangle = I_0^{*}(f\bar{u}/z)I_0(f\bar{u}/z)\langle S(\bar{u})S^{*}(\bar{u})\rangle \qquad (4)$$

or equivalently

$$\langle |I(\bar{u})|^2\rangle = |I_0(f\bar{u}/z)|^2\langle |S(\bar{u})|^2\rangle. \qquad (5)$$

The squared modulus of the OTF is then approximated by observations of an unresolvable point star, and the PSD of the object is calculated by simple inverse filtering, as

$$|I_0(f\bar{u}/z)|^2 = \langle |I(\bar{u})|^2\rangle / \langle |S(\bar{u})|^2\rangle. \qquad (6)$$

It is well known that inverse filtering is often inadequate since it tends to amplify noise at those frequencies where there is little or no energy in the OTF. It has not been found necessary in our work to use other filtering strategies, however, for the following reasons. First, because it has essentially a triangular or Bartlett frequency response, the PSD of the OTF has appreciable energy in all spatial frequencies out to the absolute

frequency cutoff imposed by the diffraction-limit of the telescope. Second, there is non-negligible noise energy in our measurements, and averaging together many noisy observations tends to put some noise energy in all the frequencies. If enough images are averaged, the inverse filtering operation is stable.

PSD estimation is presently a well-established technique in many fields, and is described in a number of texts [56-58]. By the Weiner-Khintcine theorem the PSD and the autocorrelation of a given function are a Fourier transform pair. Therefore the Labeyrie estimate is equivalently represented as the object autocorrelation. This is sufficient for characterizing centrosymmetric objects and for measuring interstar distances for binary stars. However, without knowledge of the object phase, no object reconstruction can be made in the general case.

The PSD can also be described as the zero-shift term in the autocorrelation of the Fourier transform. The Knox-Thompson estimator is similar to that of Labeyrie in that it uses terms (other than the zero-shift term) from the statistical autocorrelation of the image transform, this time to estimate phase differences. Form the product

$$\langle I(\bar{u}_1)I^*(\bar{u}_2)\rangle = I_0^*(f\bar{u}_1/z)I_0(f\bar{u}_2/z)\langle S(\bar{u}_1)S^*(\bar{u}_2)\rangle. \quad (7)$$

If $I_0(\bar{u})$ is represented in polar form

$$I_0(f\bar{u}/z) = |I_0(f\bar{u}/z)|\exp\{j\theta(f\bar{u}/z)\} \quad (8)$$

then it is easy to see that the phase differences can be

recovered from the expression

$$\frac{\langle I(\bar{u}_1)I^*(\bar{u}_2)\rangle\,|\langle S(\bar{u}_1)S^*(\bar{u}_2)\rangle|}{|\langle I(\bar{u}_1)I^*(\bar{u}_2)\rangle|\,\langle S(\bar{u}_1)S^*(\bar{u}_2)\rangle} = \exp\{j\theta(f\bar{u}_2/z)-j\theta(f\bar{u}_1/z)\}. \quad (9)$$

The limitations of this approach are derived theoretically by Knox in his thesis, assuming that the effects of atmospheric turbulence are adequately approximated by a random phase model with large variance, i.e. greater than one wavelength. (While it is known that some random amplitude variations are also introduced by the atmosphere, it is generally agreed that the phase variations have a much greater effect on image formation). Large-scale phase variations are removed by recentering the speckle images about their centroids, so that the remaining phase variation is $V(x)=\exp\{j\phi(x)\}$, where $\phi(x)$ is a stationary Gaussian random process. Knox shows that the mean OTF can be written as a product of functions describing separately the effects of the atmosphere and the telescope, and that the atmospheric 'seeing' defines the cutoff frequency of the mean OTF. His asymptotic evaluation of the autocorrelation of the OTF indicates the presence of a bias term with non-zero value out to the diffraction-limit if both $\bar{u}_1$ and $\bar{u}_2$ are zero, or if they are both large and their difference is small. The presence of this bias term for $\bar{u}_1$ and $\bar{u}_2$ small is assumed on physical grounds, since it is present in the limit whether $\bar{u}_1$ and $\bar{u}_2$ grow large or approach zero. (This intuition seems to be justified based on the results of simulations). Knox also shows that the

technique is quite insensitive to telescope aberrations. He notes that it is sensitive to noise and that for best results the distance between $\bar{u}_1$ and $\bar{u}_2$ should be kept small, so as to maximize the SNR.

III.3 Derivation of the Least-Squares Phase Integrator

Since phase estimation appears to be the primary difficulty in imaging through the atmosphere, one would like it to be robust, make maximum use of the data (in some sense), and be cheap computationally. Phase estimation by averaging a few line integrals is easy to compute, but obviously makes only limited use of the phase difference data, and as a result produces phase estimates of greater variance than might be desired. It also requires the use of arithmetic performed modulo 2π (in order that the averaging of different paths be consistent), and delivers the principal value of the phase, rather than the phase itself. A least-squares approach has the well-known advantage of satisfying a tractable objective function. In addition, it does use the data in a more 'general' way, and in practice leads to a robust estimator in noise whose statistics approximate (sometimes quite loosely) the Gaussian. It requires only conventional arithmetic (with the proviso that the Fourier transform be sampled closely enough so that the phase differences are themselves less than 2π), and it returns an 'unwrapped', or non-principal value phase estimate. The difficulty to date has been the

huge dimensionality of the problem. As in many other image processing problems, an image represented by an array of 256 by 256 discrete points expands to a linear system of equations having 65536 equations in as many unknowns. Iterative numerical schemes have previously been proposed to solve the problem [20-23]. However, equally accurate and much faster direct algorithms are available, as noted in Chapter II.3.

We have chosen an FA algorithm for the following reasons. First, since we are restricted to digital processing of the speckle images, our use of the FFT already requires the appropriate dimensioning of image arrays, and no further restriction is encountered. Second, the algorithm is nearly as fast as FACR, but is much less difficult to code. Third, it has a neat modular structure that allows its use on minicomputers with very small core-memory stores. In the following we will present the derivation of the least-squares equations and two solution algorithms. The first assumes non-periodic boundary conditions, and so is an approximation. It might prove useful should optical Fourier transform techniques be used in data processing. For our work the periodicity of the discrete Fourier transform (DFT) implies periodic boundary conditions, however, and in this case the second algorithm is exact.

III.3.1 The Normal equations

Consider a rectangular sampled image transform of dimension M*N, and denote the phase associated with point (i,j) by ϕ_{ij}. We define noisy phase differences in the two directions by

$$r_{ij}=\phi_{ij}-\phi_{i-1,j}+\varepsilon_{ij}$$

and

$$c_{ij}=\phi_{ij}-\phi_{i,j-1}+\eta_{ij}$$

where ε_{ij} and η_{ij} represent deviations from the true phase differences associated with the object. Our objective is to obtain estimates $\hat{\phi}_{ij}$ at each point in the image transform by integrating the noisy phase differences. The problem is over-determined, and we use a least-squares approach to phase estimation.

We assume that the noise statistics are shift-invariant and define the sum of squared errors by

$$\varepsilon=\sum_{i=1}^{M}\sum_{j=1}^{N}[(r_{ij}-\phi_{ij}+\phi_{i-1,j})^2+(c_{ij}-\phi_{ij}+\phi_{i,j-1})^2] \qquad (10)$$

initially assuming that ϕ_{ij} is zero outside the image array. Differentiating (10) with respect to ϕ_{ij} and equating the result to zero leads to the following equations:

$$4\phi_{ij}-\phi_{i-1,j}-\phi_{i+1,j}-\phi_{i,j-1}-\phi_{i,j+1}=r_{ij}-r_{i+1,j}+c_{ij}-c_{i,j+1}$$

$$2\leq i\leq M-1,\ 2\leq j\leq N-1 \qquad (11)$$

$$3\phi_{1j}-\phi_{2j}-\phi_{1,j-1}-\phi_{1,j+1}=c_{1j}-c_{1,j+1}-r_{2,j}$$

$$i=1,\ 2\leq j\leq N-1 \qquad (12)$$

$$3\phi_{M,j}-\phi_{M-1,j}-\phi_{M,j-1}-\phi_{M,j+1}=c_{M,j}-c_{M,j+1}+r_{M,j} \tag{13}$$

$$i=M,\ 2\leq j\leq N-1$$

$$3\phi_{i,1}-\phi_{i,2}-\phi_{i-1,1}-\phi_{i+1,1}=r_{i,1}-r_{i+1,1}-c_{i,2} \tag{14}$$

$$2\leq i\leq M-1,\ j=1$$

$$3\phi_{i,N}-\phi_{i-1,N}-\phi_{i+1,N}-\phi_{i,N-1}=r_{i,N}-r_{i+1,N}+c_{i,N} \tag{15}$$

$$2\leq i\leq M-1,\ j=N$$

$$2\phi_{11}-\phi_{21}-\phi_{12}=-r_{21}-c_{12} \tag{16}$$

$$i=j=1$$

$$2\phi_{1,N}-\phi_{2,N}-\phi_{1,N-1}=-r_{2,N}+c_{1,N} \tag{17}$$

$$i=1,\ j=N$$

$$2\phi_{M,1}-\phi_{M-1,1}-\phi_{M,2}=r_{M,1}-c_{M,2} \tag{18}$$

$$i=M,\ j=1$$

$$2\phi_{M,N}-\phi_{M-1,N}-\phi_{M,N-1}-r_{M,N}+c_{M,N} \tag{19}$$

$$1=M,\ j=N$$

Equation (11) is valid at an interior point of the image transform array, (12)-(15) at the edges, and (16)-(19) at the corners. However, if the phase array is assumed to be periodic, as it is in our work with the speckle images because of the periodicity of the DFT, then (11) is valid for all points $1 \leq i \leq M$, $1 \leq j \leq n$, provided i-1 and j-1 are taken modulo M and N.

The equations and the algorithms for solving them in the periodic and aperiodic cases differ slightly, and the algorithms will be described separately. In both cases, the equations to be solved are analogous to those which arise in solving Poisson's equation on a rectangle using the standard five-point finite difference approximation [24-28]. In particular, we discuss two variants of an FFT-based algorithm described in [28].

III.3.2 The Aperiodic case

We first define

$$v_{ij} = r_{ij} - r_{i+1,j} + c_{ij} - c_{i,j+1} \tag{20}$$

where r_{ij} and c_{ij} are noisy phase differences. We then consider the v_{ij} to be arranged in a block column vector, with v_{ij} being the jth element of the ith block. The phase elements ϕ_{ij} are ordered in an identical manner. For large M and N (the usual case in image processing), (11)-(19) can be approximated by the system of matrix equations

$$A\phi = v \tag{21}$$

where the coefficient matrix A has the block form

$$A=\begin{bmatrix} A_o & -I & & & \\ -I & A_o & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & A_o & -I \\ & & & -I & A_o \end{bmatrix} \tag{22}$$

with

$$A_o=\begin{bmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{bmatrix}. \tag{23}$$

A is block tri-diagonal of block dimension M*N, I is the N*N identity matrix, and A_o is an N*N tri-diagonal matrix.

The block matrix can be written in the alternative form

$$A_o\phi_1-\phi_2=v_1 \qquad \ell=1, \tag{24}$$

$$-\phi_{\ell-1}+A_o\phi_\ell-\phi_{\ell+1}=v_\ell \qquad 2\leq\ell\leq M-1, \tag{25}$$

and

$$-\phi_{M-1}+A_o\phi_M=v_M \qquad \ell=M. \tag{26}$$

Since A_o is symmetric, there exists an orthogonal matrix Q with Q'Q=I, where the prime denotes matrix transpose, and

$$Q'A_oQ=D_o. \tag{27}$$

As is well known, the matrix Q has as its columns the normalized eigenvectors of A_o, so

$$q_{ij}=\left(\frac{2}{N+1}\right)^{1/2}\sin\left(\frac{ij\pi}{N+1}\right) \qquad i,j=1,2,\ldots N \tag{28}$$

Note that for this case $Q=Q'$. D_o is a diagonal matrix whose diagonal elements are the eigenvalues of A_o:

$$D_{o_{ii}} = 4-2\cos\left(\frac{i\pi}{N+1}\right) \tag{29}$$

Using (27), we can write (24)-(26) in the form

$$D_o\tilde{\phi}_1 - \tilde{\phi}_2 = \tilde{v}_1 \tag{30}$$

$$-\tilde{\phi}_{\ell-1} + D_o\tilde{\phi}_\ell - \tilde{\phi}_{\ell+1} = \tilde{v}_\ell \tag{31}$$

and

$$-\tilde{\phi}_{M-1} + D_o\tilde{\phi}_M = \tilde{v}_M \tag{32}$$

where

$$\tilde{\phi}_i = Q'\phi_i \tag{33}$$

and

$$\tilde{v}_i = Q'v_i \tag{34}$$

Alternatively,

$$\begin{bmatrix} D_o & -I & & & \\ -I & D_o & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & D_o & -I \\ & & & -I & D_o \end{bmatrix} \tilde{\phi} = \tilde{v} \tag{35}$$

Permuting the rows and columns of this coefficient matrix by grouping the ℓth equation in each block together results in the alternative coefficient matrix

$$\Lambda = \begin{bmatrix} \Lambda_1 & & & \\ & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_M \end{bmatrix} \tag{36}$$

where

$$\Lambda_\ell = \begin{bmatrix} D_{o_{\ell\ell}} & -1 & & & \\ -1 & D_{o_{\ell\ell}} & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & D_{o_{\ell\ell}} & -1 \\ & & & -1 & D_{o_{\ell\ell}} \end{bmatrix} \qquad (37)$$

In other words, the original block tri-diagonal system can be decomposed into M decoupled tri-diagonal systems, each of which can be quickly solved using Gauss elimination. Having solved these decoupled systems, we then undo the permutation described above and premultiply the result by Q to obtain the solution to the original system.

We see from (28) the the diagonalization process essentially amounts to calculating a DFT. If N+1 and M+1 are highly composite, in particular if they are powers of 2, this computation can be performed rapidly using the fast Fourier transform (FFT) algorithm. To accomplish this, we first define for each k

$$\mathrm{Re}\{t_\ell\} = \begin{cases} 0 & \ell = 0 \\ v_{k\ell} & \ell = 1,2,\ldots N \\ 0 & \ell = N+1,\ldots 2(N+1)-1 \end{cases} \qquad (38)$$

and

$$\mathrm{Im}\{t_\ell\} = 0 \qquad \ell = 0,1,\ldots 2(N+1)-1 \qquad (39)$$

Then

$$\tilde{v}_{k\ell} = \left(\frac{2}{N+1}\right)^{1/2} \mathrm{Im}\{\mathrm{DFT}(t_\ell)\} \qquad \ell = 1,2,\ldots N \tag{40}$$

where the DFT is defined by

$$\mathrm{DFT}(t_\ell) = \sum_{\ell=0}^{2(N+1)-1} t_\ell \exp\left[\frac{jk\ell 2\pi}{2(N+1)}\right] \tag{41}$$

Storage requirements and computation time can be further reduced by taking advantage of the even and odd symmetry properties of the DFT. Because they are well known, we do not discuss them here, although our programs do incorporate them.

III.3.3 The Periodic case

The strategy in the periodic case is similar to that outlined above. In this case, the coefficient matrix is the block circulant matrix

$$C = \begin{bmatrix} C_o & -I & 0 & \ldots & -I \\ -I & C_o & -I & \ldots & 0 \\ & & \vdots & & \\ -I & 0 & \ldots & -I & C_o \end{bmatrix} \tag{42}$$

where

$$C_o = \begin{bmatrix} 4 & -1 & 0 & \ldots & -1 \\ -1 & 4 & -1 & \ldots & 0 \\ & & \vdots & & \\ -1 & 0 & \ldots & -1 & 4 \end{bmatrix} \tag{43}$$

For this case, Q=Q'. Q has as two of its columns $(1/N)^{1/2}(1,1,\ldots,1)'$ and $(1/N)^{1/2}(1,-1,1,\ldots,-1)'$, with corresponding simple eigenvalues 2 and 6, respectively. In addition, there are (N-2/2) double eigenvalues

$$\gamma_\ell = 4-2\cos(2\pi\ell/N) \qquad \ell=1,\ldots(N-2)/2 \tag{44}$$

each associated with a pair of eigenvectors

$$\epsilon_k^{(\ell)} = \left(\frac{2}{N}\right)^{1/2} \sin\left(\frac{k\ell 2\pi}{N}\right) \tag{45}$$

and

$$\eta_k^{(\ell)} = \left(\frac{2}{N}\right)^{1/2} \cos\left(\frac{k\ell 2\pi}{N}\right) \tag{46}$$

Diagonalization is again carried out using the FFT. In this case, we require N to be highly composite, preferably a power of two. After diagonalization and permutation as in the aperiodic algorithm, the decoupled block coefficient matrices become

$$\Gamma_\ell = \begin{bmatrix} \gamma_\ell & -1 & 0 & \ldots & & -1 \\ -1 & \gamma_\ell & -1 & 0 & \ldots & 0 \\ 0 & -1 & \gamma_\ell & -1 & 0 \ldots & 0 \\ & & \vdots & & & \\ -1 & 0 & \ldots & & -1 & \gamma_\ell \end{bmatrix} \tag{47}$$

where

$$\gamma_\ell = 4-2\cos(2\pi\ell/N), \qquad \ell=0,1,\ldots N/2 \tag{48}$$

As described above for the aperiodic case, these decoupled systems can now be solved separately. The permutation and

diagonalization operations are then reversed to obtain the solution to the original system of equations.

There are some important differences between the periodic and aperiodic cases. First, Γ_o is singular, reflecting the singularity of C. If the first row and column of Γ_o are deleted, a solution is obtained which satisfies the original equations and which differs from any other solution by an additive constant. This ambiguity can be resolved in our case by noting that the phase at the origin must be zero.

Second, the remaining blocks of coefficients are not tri-diagonal since they have non-zero corner elements. This situation can be dealt with using the Sherman-Morrison formula [28], which states that if $A=B + uv'$, where u and v are column vectors, then

$$A^{-1}=B^{-1}+B^{-1}u(1 + v'B^{-1}u)^{-1}v'B^{-1}$$

If v contains only one nonzero element, the correction is quickly obtained. Since most of the eigenvalues are of order two, two blocks can usually be solved at once, thereby compensating somewhat for the additional computation involved in the Sherman-Morrison formula. A further saving is realized from the increased efficiency of the FFT in diagonalizing the periodic structure.

III.4 Algorithm performance

A number of tests were run with both the periodic and aperiodic integration algorithms to determine their

accuracy. The first tests related to the propagation of numerical errors irrespective of any measurement errors in the data. The other tests examined the propagation of measurement error in the integrated solution.

The first tests were as follows. An appropriately dimensioned file was filled with Gaussian distributed random noise of standard deviation 1. The discrete Poisson differential operator produced another file, which upon integration was compared to the first. Typical results for the algorithms are summarized in Tables 1 and 2.

Integration time for the 255*255 aperiodic array was about 200 seconds on our single user PDP-10 computer. (It has floating point add and multiply times of about 5 us and 11 us, respectively). Approximately 175 seconds were spent on the FFT computations. Because the periodic algorithm is better suited to an FFT implementation, the 256*256 problem required only 135 seconds, of which approximately 100 seconds were spent on the FFT. Corrections by the Sherman-Morrison formula required only about 10 seconds.

Previous researchers have indicated the stability of the growth of the mean squared error (MSE) in the integrated solution due to errors in estimating the phase differences when the least-squares approach is used. Both Fried and Hudgin estimated the growth as logarithmic with increasing array size. Based on simulations with small arrays, Fried found that a good fit to the mean square

Table 1. Error in the estimated phase due to numerical errors when using the aperiodic integrated algorithm.

Array Size	Maximum error	SSE Norm
15x15	$.33 \times 10^{-6}$	$.14 \times 10^{-13}$
31x31	$.48 \times 10^{-6}$	$.18 \times 10^{-13}$
63x63	$.9 \times 10^{-6}$	$.18 \times 10^{-12}$
127x127	$.8 \times 10^{-5}$	$.37 \times 10^{-12}$
255x255	$.23 \times 10^{-5}$	$.80 \times 10^{-11}$

Table 2. Error in the estimated phase due to numerical errors when using the periodic integrated algorithm.

Array Size	Maximum Error	SSE Norm
16x16	$.21 \times 10^{-6}$	$.73 \times 10^{-14}$
32x32	$.66 \times 10^{-6}$	$.98 \times 10^{-13}$
64x64	$.98 \times 10^{-6}$	$.83 \times 10^{-12}$
128x128	$.20 \times 10^{-5}$	$.41 \times 10^{-12}$
256x256	$.48 \times 10^{-5}$	$.46 \times 10^{-11}$

reconstruction error could be made with the function

$$e = \sigma_{pd}^2[.3205\ \ln(N) + .6558] \quad (49)$$

where σ_{pd}^2 is the variance of the errors made in estimating the phase differences. Similarly, Hudgin obtained the function

$$e = \sigma_{pd}^2[.103\ \ln(N) + .561] \quad (50)$$

Hunt derives an expression for the error based on periodic boundary conditions, which he gives as

$$e = (\sigma_{pd}^2/N^2)\{\sum_{ij}[2\sin(i\pi/N)/4-2\cos(2\pi i/N)-2\cos(2\pi j/N)]$$
$$+ [2\sin(j\pi/N)/(4-2\cos(2\pi i/N)-2\cos(2\pi j/N))]^2\}. \quad (51)$$

He shows that the mean square error in the integrated phase estimate predicted by his trigonometric expression has the form of the empirically derived logarithmic expressions of Fried and Hudgin. He also notes that upon plotting the results there is a distinct offset between his predictions and the results of the other two. This offset is ascribed to the different ways in which their algorithms actually handled the boundary conditions. A plot of these three curves together with typical solutions from our periodic algorithm is found in Fig. 1.

The stability of the reconstructions with increasing array size is encouraging. Note also that the MSE of the reconstructed phase estimates is actually less than the variance of the phase difference measurement error for array sizes up to 256*256. Because of the periodicity of the DFT Hunt's assumption of periodic boundary conditions

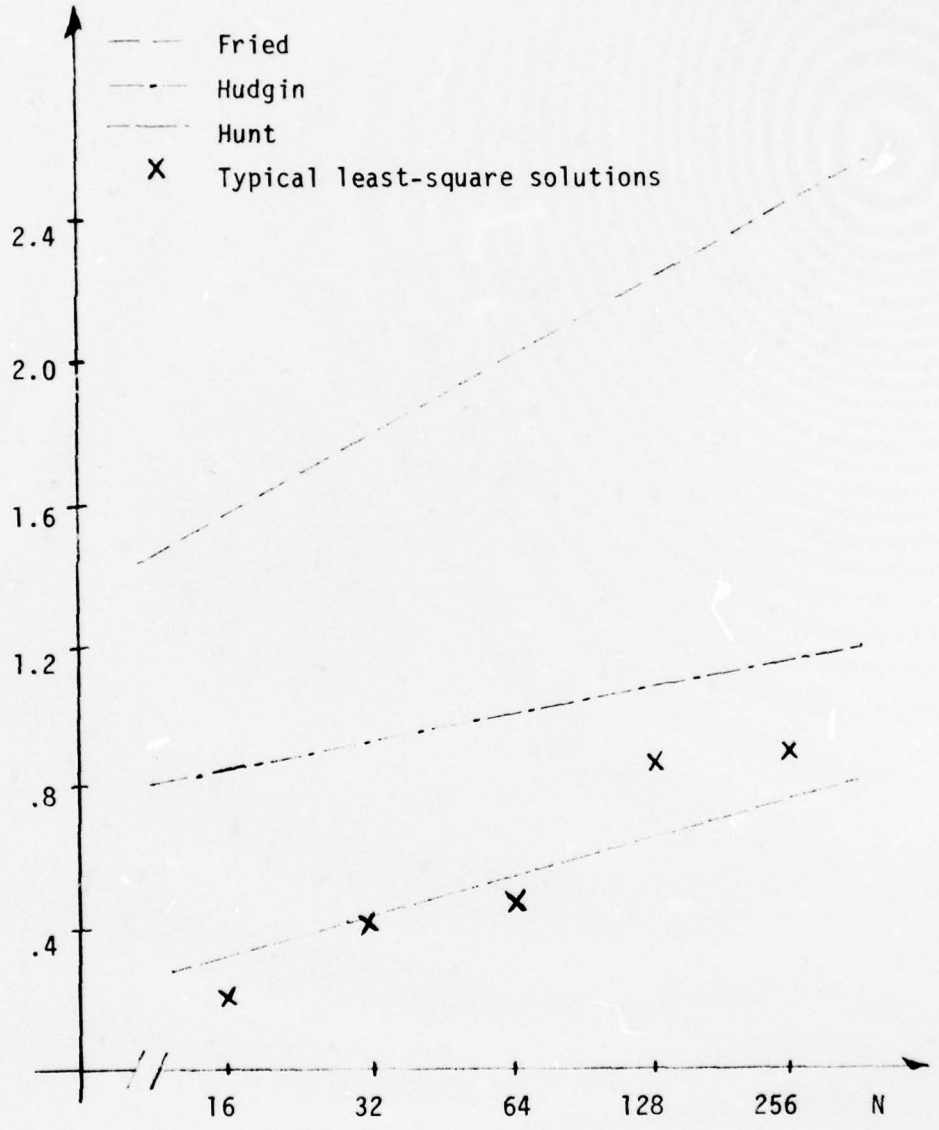


Fig. 1. Growth of the ratio of mean square error in the integrated phase estimate to the variance of the error in measuring phase differences as a function of array size N.

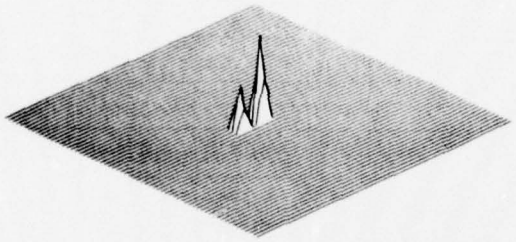
is exactly satisfied, and our results agree quite well with his analytically derived result.

III.5 Phase Estimation on Simulated Data

Simulations were used extensively to evaluate the effectiveness of the various estimation procedures discussed in this research. The method of generating synthetic data is described below. Note that the resulting synthetic data base is common to all the simulations in this dissertation.

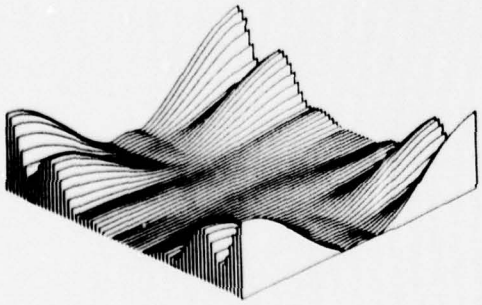
A synthetic binary star was constructed of a separable triangle of height 1 and a shifted and scaled version of itself; it is shown in Fig. 2a. It is defined on a 64*64 array and has a region of support of 18 pixels. Its associated magnitude and phase spectra are shown in Figs. 2b and 2c, respectively. (Note that the illustrations of the frequency spectrum are represented in the base band with the zero frequency in the upper left hand corner. They are also transposed relative to the intensity object, to save processing time. This convention will hold throughout the dissertation). This object is blurred by a digital process imitating the effects of atmospheric turbulence. We assume that the distortions can be described by a Gaussian phase model, and that these phase distortions are the principal cause of image degradation. These atmospheric effects were simulated by randomizing the phase of the pupil function of the

MIN 0000000 (0000000)
MAX= 1000000e1 (1000000e1)
AVGE= 1526768)



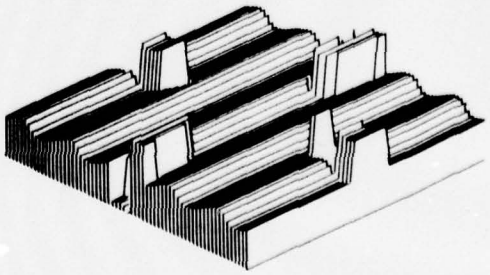
a) Intensity Object

MIN= 0000000 (0000000)
MAX= 6257359e1 (6257359e1)
AVGE= 1078611e1



b) Object Magnitude Spectrum

MIN=- 3141593e1 (- 3141593e1)
MAX= 3141593e1 (3141593e1)
AVGE=- 2032018e-1



c) Object Phase Spectrum

Fig. 2 Synthetic Double Star Used in all Simulations

incoherent imaging system.

A series of arrays of Gaussian random noise variables with unit variance was generated from a random process which was zero mean, unit variance, and uncorrelated. Correlated noise was produced by filtering these arrays by the technique described in Knox [15], with the correlation length corresponding to a turbulence cell size of 11 inches across a 56-inch telescope. Fig. 3 shows one of these random number arrays. The point-spread function of the digital correlation filter is shown in Fig. 4. A correlated noise array is illustrated in Fig. 5. Each correlated noise array was used as the phase $\theta(k)$ of a complex pupil function whose magnitude equaled the aperture function $A(\bar{x})$ of Fig. 6. The pupil functions were inverse Fourier transformed and the squared modulus formed yielding a set of 45 different point-spread functions. One of these is shown in Fig. 7.

An independent set of 45 point-spread functions was similarly generated. These were convolved with the double star image of Fig. 2 to yield the set of 45 blurred images $i(\bar{x})$, one of which is shown in Fig. 8. These blurred double star images represent the short-exposure photographs which are then input to the Knox-Thompson procedure.

To examine the sensitivity of this technique to sensor noise, we performed a series of simulations with different levels of sensor noise present. The sensor noise model is

MIN= -.3657204@1 (-.3657204@1)
MAX= .3286373@1 (.3286373@1)
AVGE= .3401955@-2

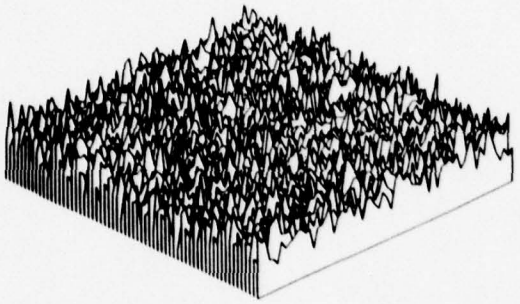


Fig. 3 Random numbers (to be used as the Phase of the Instantaneous OTF).

MIN= -.2439319@-1 (-.2439319@-1)
MAX= .9862156@2 (.9862156@2)
AVGE= .7978798

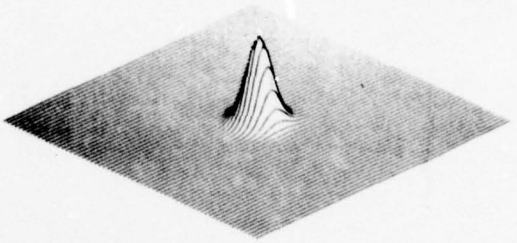


Fig. 4 Gaussian Spatial Phase Correlation Filter

MIN= -.1590778@4 (-.1590778@4)
MAX= .1059941@4 (.1059941@4)
AVGE= -.1098180@2

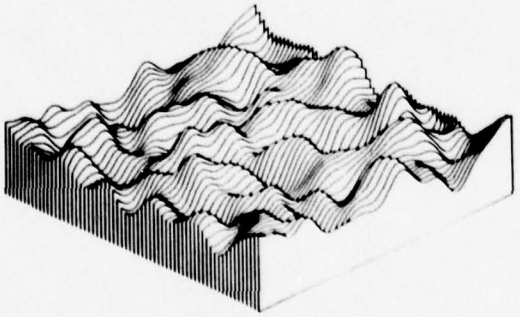


Fig. 5 Spatially Correlated Phase of Instantaneous OTF.

MIN= .0000000 (.0000000)
MAX= .1000000@1 (.1000000@1)
AVGE= .2085808

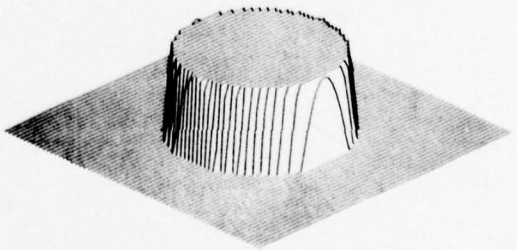


Fig. 6 Aperture Function of Telescope.

MIN= 49116918-2 (49116918-2)
MAX= 257360505 (257360505)
AVGE= 8387977E3

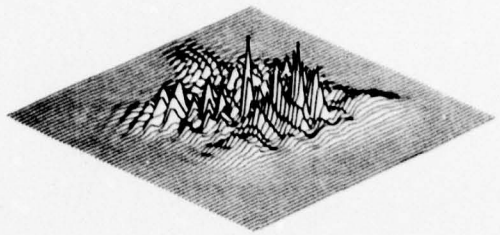


Fig. 7 Simulated Point Spread Function

MIN= 473135701 (473135701)
MAX= 677648405 (677648405)
AVGE= 524865984

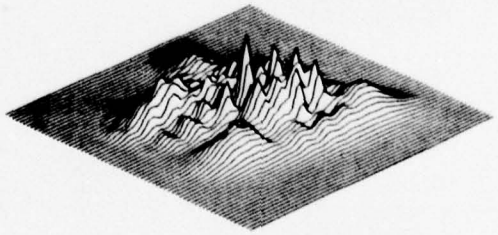


Fig. 8 A Simulated Blurred Double Star Image

based on a semi-classical approach to photon detection. In this model, the number of photoelectrons released from a small region of area A centered at a point (x,y) in the image plane is taken to be a Poisson random variable with mean

$$\lambda=[\eta I(x,y)/h\bar{\nu}]\tau\Delta A.$$

In this expression, η is the quantum efficiency of the photodetector, h is Planck's constant, $\bar{\nu}$ is the mean optical frequency, τ is the integration (exposure) time, and I(x,y) is the image intensity at point (x,y).

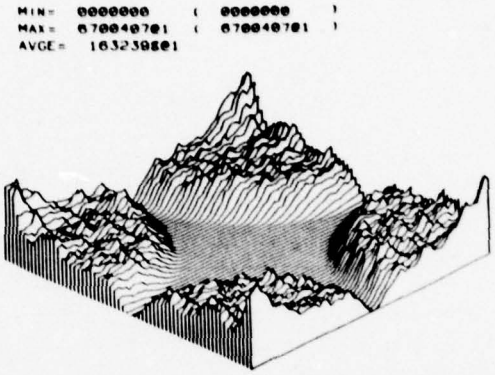
It is well known that both the mean and the variance of the above Poisson random variable are equal to λ. Thus, the mean photoelectron current produced by a detector at point (x,y) is proportional to the image intensity at that point, but so are the fluctuations in that current. If we define the SNR of the image at point (x,y) as the ratio of the square of the mean current to the variance of the current, we find that this SNR is just λ. We can vary λ, and therefore the SNR, by varying τ (the integration time) or ΔA (the area over which the image is averaged).

An additional simplification results if we assume that the number of photoelectrons is large. In this case, the photoelectric current will be approximately Gaussian with a mean equal to its variance. We made this assumption in our simulations. The SNR which we ascribe to a given simulated image is simply the maximum of the point-by-point ratios of

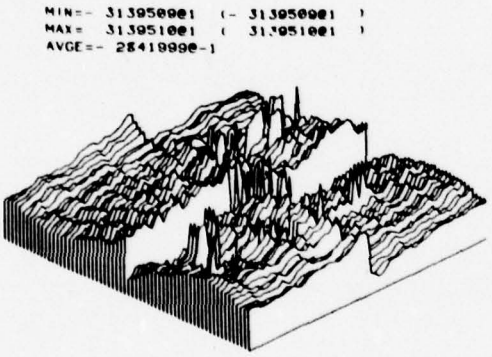
squared mean to variance as described above. This definition is somewhat arbitrary, of course, but this is not a serious problem since we are interested primarily in relative performance as we vary the noise.

Results of the Knox-Thompson restoration procedure, comparing both line integration and least-squares integration for various SNR as defined above, are shown in Figs. 9a through 9y. Although we would like to quantify the results in some perceptually significant manner, no general perceptual distance metric has yet been defined for images, much less for the way in which phase errors might affect image formation. We default then to SSE measures, and simply include illustrations of the various estimates and associated reconstructions. The WSSE measure in the captions of the phase estimates refers to a weighted SSE measure where the weighting function is the magnitude of the object Fourier transform. The phase error measurements reflect the modulo 2π nature of the phase. For reference the object energy is 2.929 units and the sum squared total, or 'energy' of the phase, is 3860.642 units.

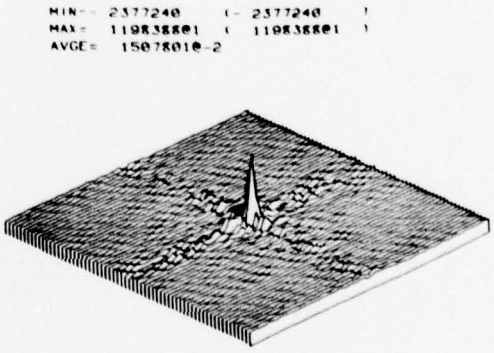
All the reconstructions were based on the Labeyrie magnitude spectrum estimate. To minimize artifacts in the reconstructed image, these magnitude estimates were digitally low-pass filtered with a radially symmetric function which was constant to 25 cycles/picture, and then tapered to zero at 32 cycles/picture by means of a cubic



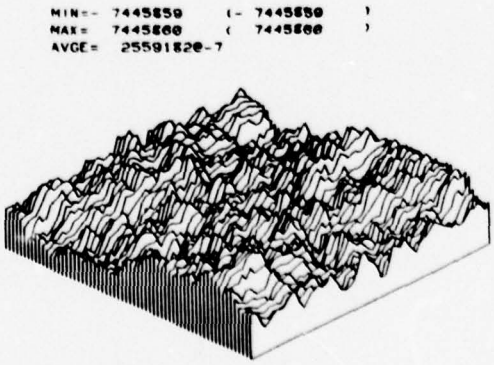
a)Magnitude Spectrum
SSE = 4326.848



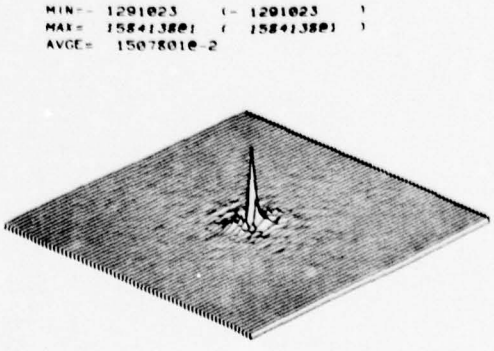
b) Line-Integral Phase
WSSE = 1232.070



c) Line-Integral Object
SSE = 1.960/.682

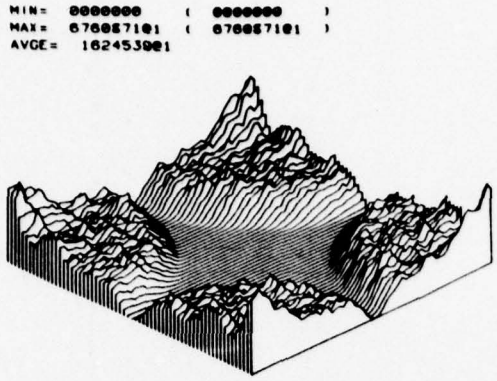


d) Least-Square Phase
WSSE = 463.461

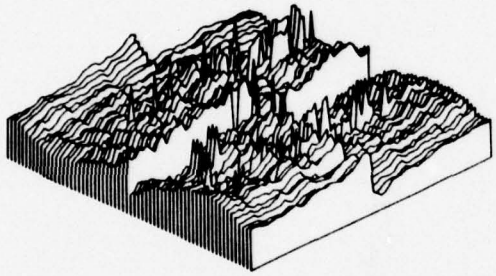


e) Least-Square Object
SSE = 1.423/.948

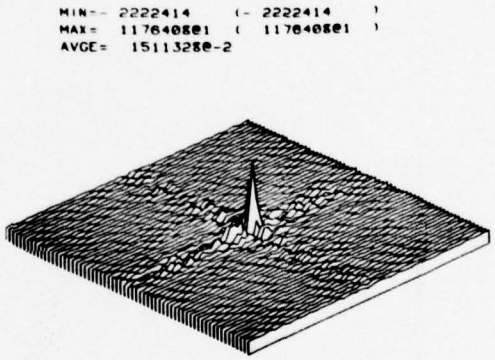
Fig. 9(a-e) Comparison of Line-Integral and Least-Square Reconstruction at 5 dB SNR



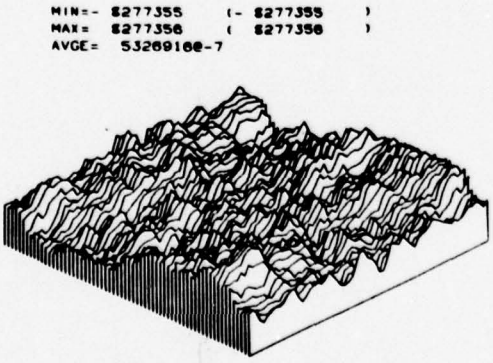
f) Magnitude Spectrum
SSE = 4038.903



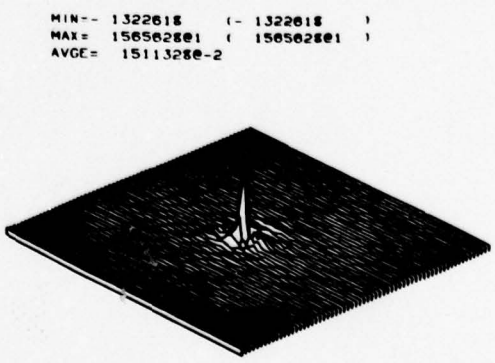
g) Line-Integral Phase
WSSE = 1420.542



h) Line-Integral Object
SSE = 2.024/.663

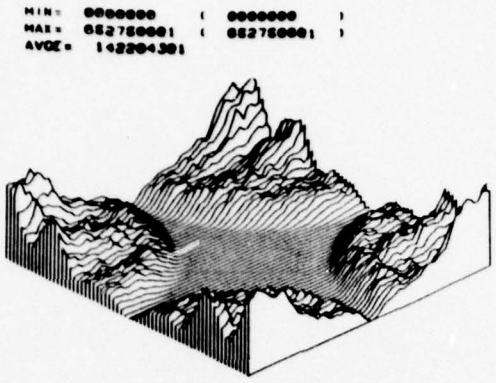


i) Least-Square Phase
WSSE = 471.632

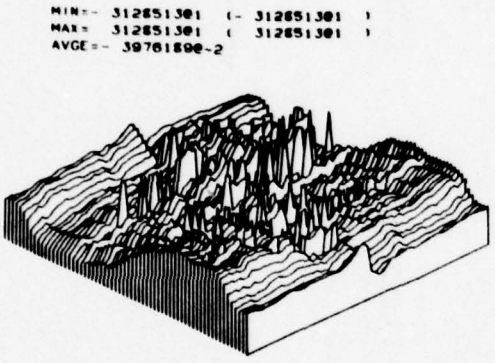


j) Least-Square Object
SSE = 1.369/.903

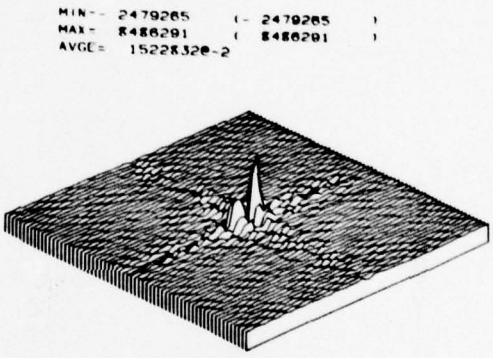
Fig. 9(f-j) Comparison of Line-Integral and Least-Square Reconstruction at 10 dB SNR



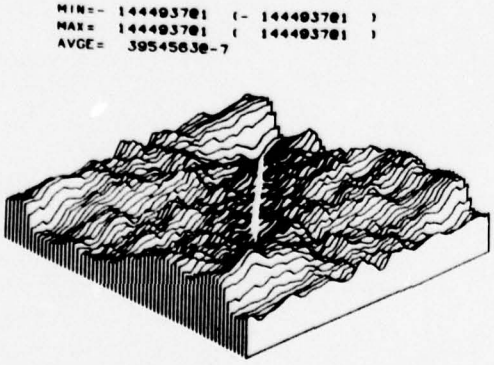
k) Magnitude Spectrum
SSE = 1648.057



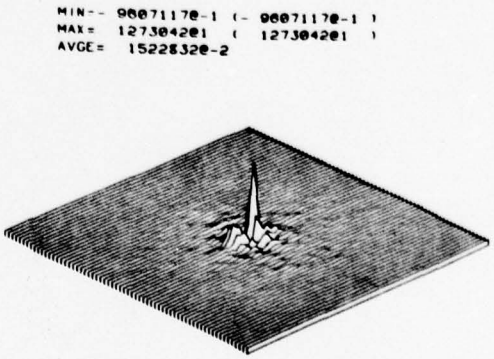
l) Line-Integral Phase
WSSE = 3765.697



m) Line-Integral Object
SSE = 2.811/1.411

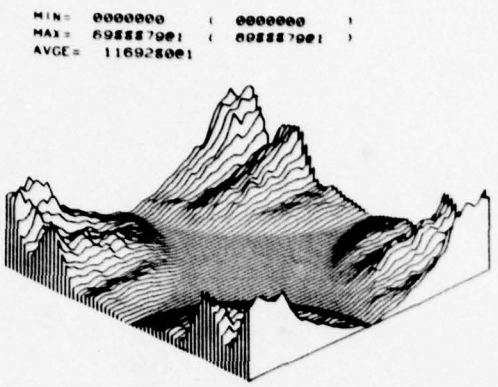


n) Least-Square Phase
WSSE = 729.081

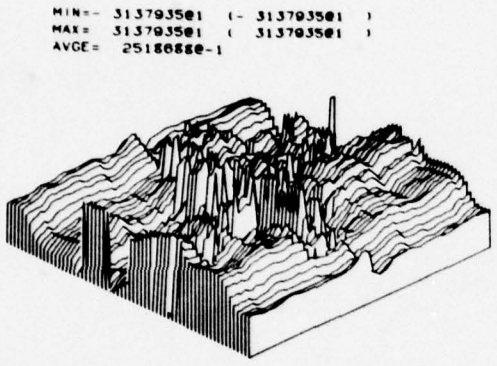


o) Least-Square Object
SSE = .989/.608

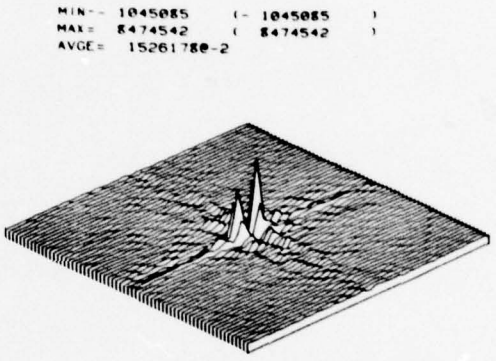
Fig. 9(k-o) Comparison of Line-Integral and Least-Square Reconstruction at 20 dB SNR



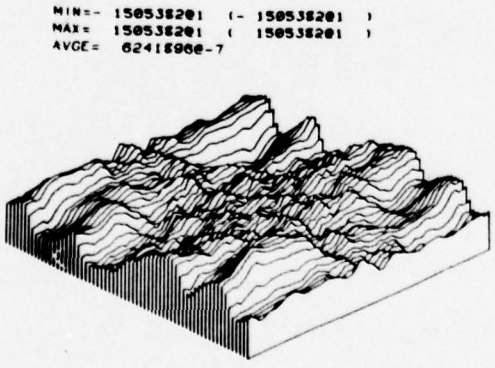
p) Magnitude Spectrum
SSE = 403.047



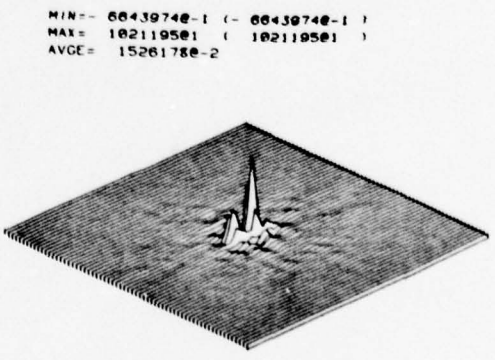
q) Line-Integral Phase
WSSE = 5467.787



r) Line-Integral Object
SSE = 2.798/1.547

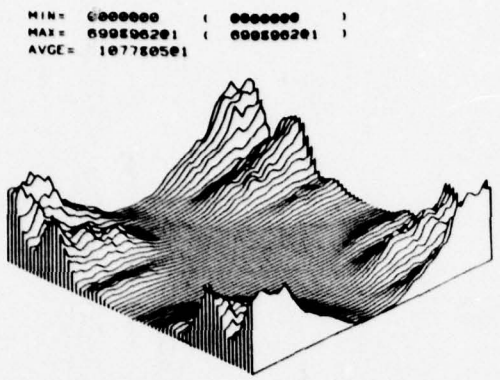


s) Least-Square Phase
WSSE = 777.555

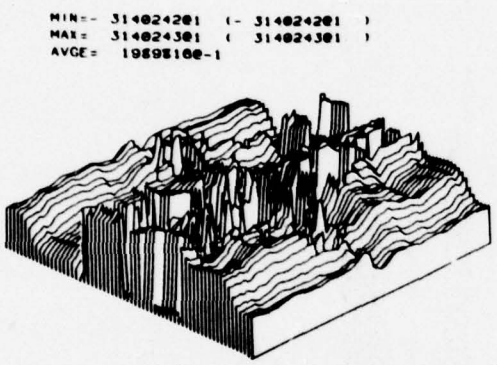


t) Least-Square Object
SSE = .514/.277

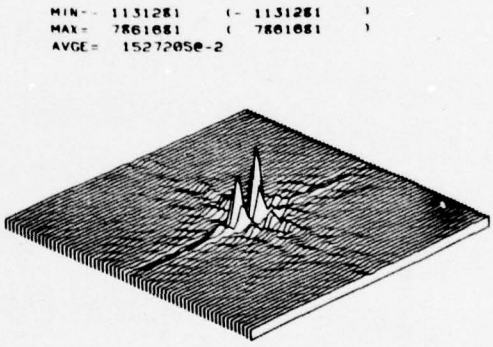
Fig. 9(p-t) Comparison of Line-Integral and Least-Square Reconstruction at 30 dB SNR



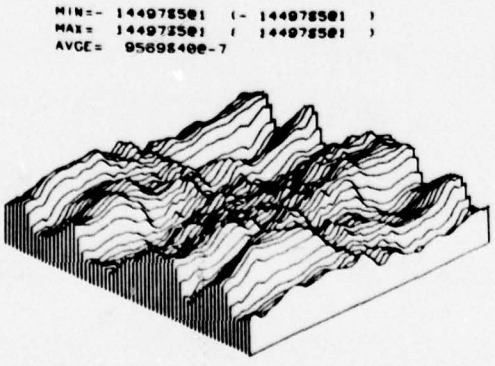
u) Magnitude Spectrum
SSE = 275.568



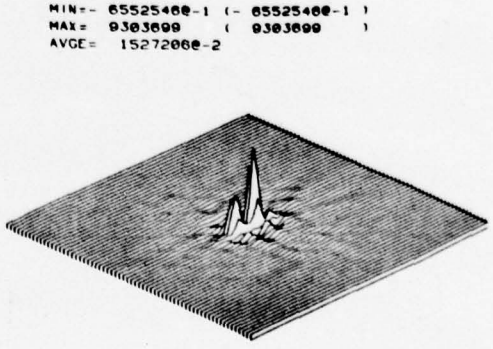
v) Line-Integral Phase
WSEE = 5953.354



w) Line-Integral Object
SEE = 2.789/1.443



x) Least-Square Phase
WSSE = 772.358



y) Least-Square Object
SSE = .435/.220

Fig. 9(u-y) Comparison of Line-Integral and Least-Square Reconstruction at 40 dB SNR

spline function.

At a SNR of 5 dB, neither method gives a good indication of a second star. The least-squares method is more localized, however, and does not show the presence of artifacts along the image coordinate axes. At 10 dB the least-squares restoration still shows only one star, while the line integral restoration hints at the possibility of two stars. At 20 dB both reconstructions indicate clearly the second star. The line integral version has more correctly estimated the relative amplitudes of the stars, but also has many more artifacts. At 30 dB and 40 dB both methods have isolated the two stars. The least-squares versions are more nearly correct in relative amplitude and are less noisy. (The SSE measures are somewhat confusing on the phase, where they get better as the estimate gets worse. Note also the two SSE measures on the reconstructed objects. The first refers to the error over the entire field of view, the second to the error within the region of support of the star. For the line integral object reconstructions, both measures behave erratically).

Both types of restorations have regions of negative light intensity. These nonphysical artifacts are a result of misestimation in the Fourier domain, and are quite common in deconvolution problems. They are called super-blacks in the literature [55]. In every case the peak negative light intensity is greater in magnitude in

the line integral restorations than in the least-squares restoration, and has a magnitude of 10 to 20 percent of the positive light-intensity peak. The restorations also tend to be smeared, and occupy a larger area in the field of view than the object. Less distracting is the fact that the restorations are complex, instead of real. However, the magnitude of the imaginary part is generally about 2 orders of magnitude less than the real part, and is entirely due to the phase estimate not having perfectly odd symmetry. In an effort to further improve the reconstructions, a number of modifications were tried.

The first modification was the inclusion of phase difference information derived from the second-shift term of the ACF of the image transforms. The necessary modification of the periodic least-squares integration algorithm is straightforward. We experimented with an integrator which weighted differently the phase differences from the one- and two- shift terms, but found that the WSSE of the resulting estimates was a monotonically increasing function of the weight applied to the second-shift data. This is consistent with Knox's analysis and recommendation that only one-shift data be used.

We also experimented with the use of spatial domain windows applied to each speckle photograph. This is a common technique in spectral estimation for segmenting long data sequences into manageably short lengths. The effect

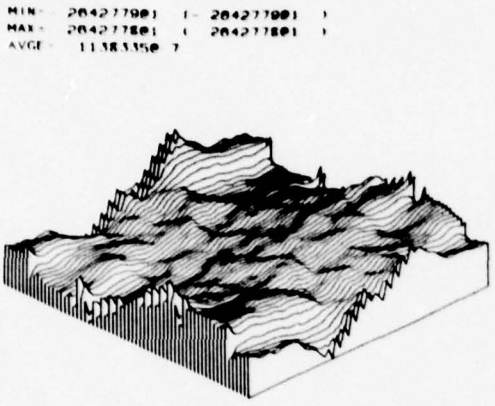
is of course to convolve the transform of each speckle image with the transform of the window function before the averaging. The effects on the bias and variance of the subsequent PSD estimates are found in Jenkins' and Watts' text [56]. We wanted to improve the effective SNR by restricting ourselves to the brightest part of each speckle image. We also found in working with the real data that the Knox-Thompson estimator seemed much less accurate than our simulations would indicate. We hypothesized that it might be quite sensitive to the violation of the isoplanatic assumption, so we wanted to restrict the field-of-view to more nearly approximate isoplanicity.

Although the effects of windowing on PSD estimates have been examined by many researchers, the effects on the Knox-Thompson estimator have not been so catalogued. Therefore, we examined the effects of spatial windowing on the phase difference estimates by the same simulation techniques previously described. A series of blurred stars was created and processed with each one of three different spatial windowing options to produce three different reconstructions. In the first case each speckle photograph was multiplied by a Fourier (rectangular) window of dimension 32*32. The second reconstruction used a radially symmetric window made by rotating a Hamming window of radius 20 about its central axis. The third reconstruction used no spatial window, and is that of Figs. 9o and 9p.

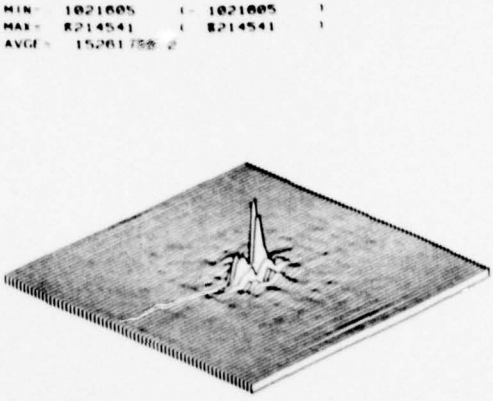
The example illustrated in Fig. 10 was performed with a nominal SNR of 30 dB. In all three cases the same magnitude spectrum estimate (that of the unwindowed version) was used, so as to isolate the effects on the phase estimation. As before, the SSE measures are found in the figure captions. Note that the use of a Fourier window degrades the results, although the Hamming reconstruction is actually improved, at least in terms of SSE. This is not unexpected behavior. Similar behavior was observed in all other test cases and in processing the real data.

III.6 Phase Refinement

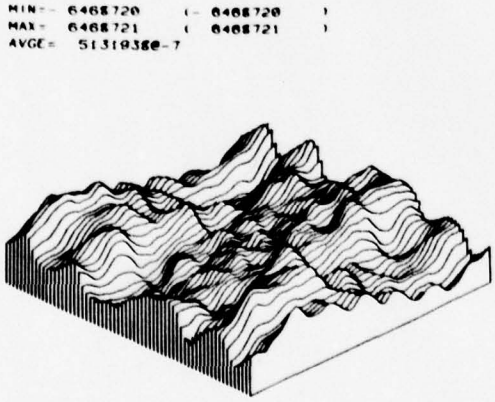
As mentioned in the literature review, the Gerchberg-Saxton algorithm was created to solve the phase problem of electron microscopy, where one has knowledge of the magnitude of both a function and its Fourier transform but no knowledge of the phase in either domain. A flow chart of the algorithm is found in Fig. 11. The algorithm is almost directly applicable to the problem of phase estimation where one has knowledge of only the magnitude spectrum and the region of support of a real function and no knowledge of its phase spectrum. However, this relaxation of the required a priori knowledge introduces a serious non-uniqueness problem in the reconstructions, which has been demonstrated empirically by Feinup [40]. (He has some experimental evidence to indicate that the problem may not be as severe for complicated two



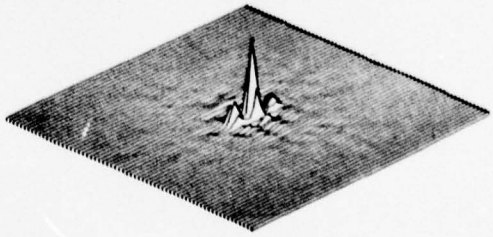
a) Fourier Window
WSSE = 1832.647



b) Fourier Window
SSE = 1.562/.938



c) Hamming Window
WSSE = 441.695



d) Hamming Window
SSE = .364/.187

Fig. 10 Comparison of the Effect of Different Spatial Windows on the Least-Square Reconstructions at 30 dB SNR

dimensional images, however). In this work we have sought to side-step the possible problem of non-uniqueness by the use of the least-square Knox-Thompson estimate as an initial approximation to the correct phase spectrum. The hope is that it is 'close enough' in some sense to the true phase spectrum to allow the algorithm to converge to the true phase.

The new algorithm is presented in flow chart form in Fig. 12. The only changes are that instead of correcting the magnitude in the spatial domain, one imposes the conditions of realness, non-negativity, and a bounded region of support on the reconstruction. To prove that the correction energy must decrease only a slight variation need be made to the geometric proof originally offered by Gerchberg and Saxton. Consider the operation of the algorithm on two discrete points, one in the spatial domain and the other in the frequency domain, as in Fig. 13. Define the squared error e as the sum squared difference between the original magnitudes at each discrete frequency and the magnitudes calculated by the algorithm at corresponding points. When e is zero, the algorithm has converged. Now, in the frequency domain, let the reference magnitude at our arbitrary point be represented by the measured magnitude f_2. Upon inverse Fourier transformation, a vector f_1 is produced at the arbitrary point in the spatial domain. If f_1 is in the region of

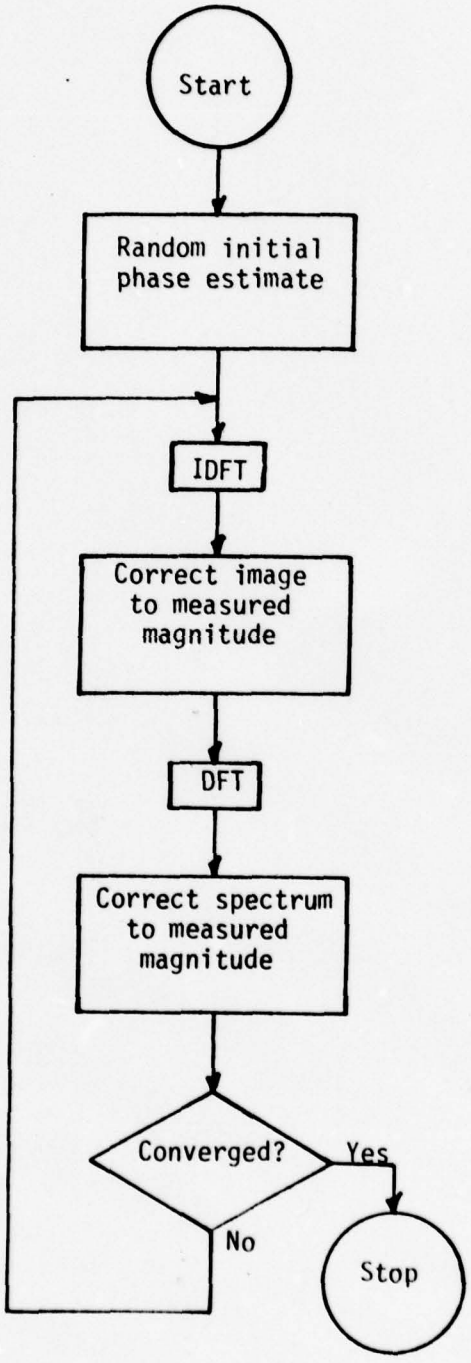


Fig. 11. Flow chart of Gerchberg-Saxton algorithm used in electron microscopy.

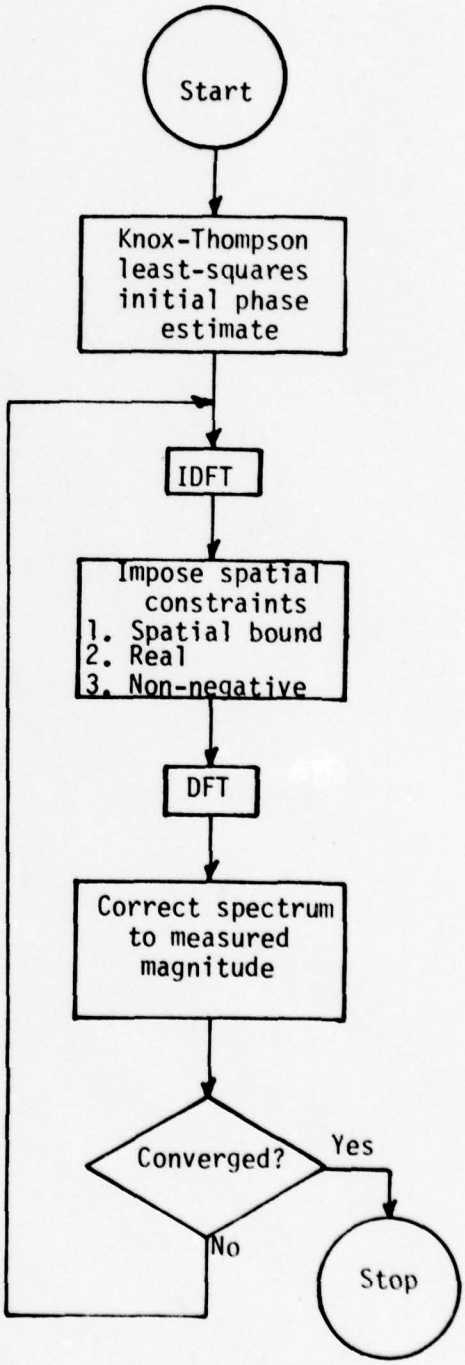


Fig. 12. Flow chart of modified Gerchberg-Saxton algorithm used for phase refinement.

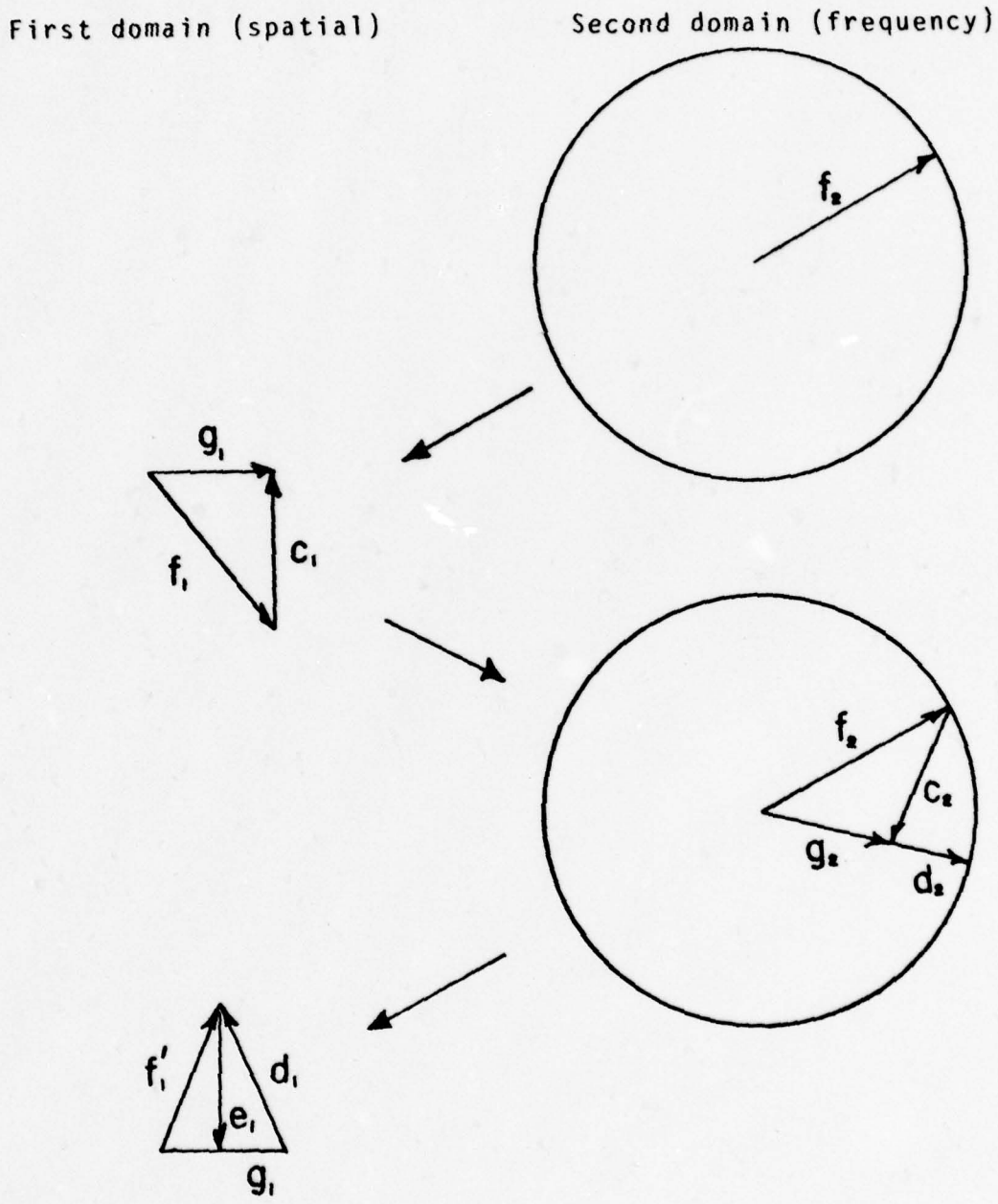


Fig. 13. A geometric proof of the convergence of the Gerchberg-Saxton algorithm when used for phase refinement.

support and if its real part is non-negative the real part is preserved. Otherwise f_1 is set to zero by the addition of a colinear vector of equal magnitude and opposite direction. All points in the spatial domain are similarly corrected and the Fourier transform of the new object is taken. At our arbitrary point in the frequency domain the new vector g_2 is the sum of f_2 and a new vector c_2 produced at that frequency by the Fourier transform of the vectors c_1 at each point in the spatial domain. If necessary the magnitude of g_2 is corrected to that of f_2 by the addition of the vector d_2, which is colinear with g_2. The algorithm now iterates until convergence. It is clear from Parseval's theorem that

$$e^2 = \sum_{ij} |c_1|^2_{ij} = \frac{1}{N^2} \sum_{ij} |c_2|^2_{ij} \tag{52}$$

It is also clear from examining the phasor plots of Fig. 13 that

$$|d_2|^2_{ij} \leq |c_2|^2_{ij} \quad \forall i,j \tag{53}$$

with equality iff c_2 is colinear with f_2. Thus

$$\frac{1}{N^2} \sum_{ij} |d_2|^2_{ij} \leq \frac{1}{N^2} \sum_{ij} |c_2|^2_{ij} = \sum_{ij} |c_1|^2_{ij} \tag{54}$$

with equality iff the correction vector is colinear at every frequency. To this point we have shown that the correction energy applied in the frequency domain cannot be larger than the correction energy applied in the spatial domain just previously. Now, in the second iteration, we have in the spatial domain the vector f_1', the sum of vectors g_1 and d_1. It is geometrically obvious that

$$|e_1|^2_{ij} \leq |d_i|^2_{ij} \quad \forall i,j \tag{55}$$

Thus

$$\sum_{ij} |e_i|^2_{ij} \leq \sum_{ij} |d_1|^2_{ij} = \frac{1}{N^2} \sum_{ij} |d_2|^2_{ij} \qquad (56)$$

and the correction signal energy in the spatial domain is also less than or equal to the frequency domain correction energy applied just prior to inverse Fourier transformation. Hence the squared error must decrease or remain constant with each iteration. For the squared error to be bounded away from zero, it is required that the correction signal in the frequency domain be colinear with the estimated frequency spectrum while being entirely imaginary in the spatial domain. This condition cannot be fulfilled because the algorithm forces the spatial estimate to be real, and hence to have an even magnitude spectrum and an odd phase spectrum. To be colinear the correction signal must also have odd phase symmetry, and hence cannot be imaginary in the spatial domain. (Note that this result is stronger than that presented for the original electron microscopy problem, where certain phase symmetries in the spatial and frequency domains may bound the correction energy away from zero.) This result does not necessarily imply convergence of the correction energy to zero, nor does it imply solution uniqueness, but in our experience this proviso is of theoretical rather than practical interest.

One might wonder whether, given a phase estimate, one could effect a magnitude solution by the same algorithm.

The answer is no, because it can be shown that the correction signal energy cannot be guaranteed to decrease. Imagine again the same phasor diagrams, with the same initial signal estimates, as represented in Fig. 14. At the first return to the frequency domain, one corrects the magnitude of g_2 by adding d_2. Since d_2 must now be perpendicular to the perpendicular bisector OP, it is geometrically obvious in this case that $d_2 > c_2$. If this condition is satisfied at enough points, the correction signal energy will increase, and the algorithm may never converge. In fact, this behavior has been observed in empirical tests.

To test the phase refinement technique, the least-squares phase estimates described in the previous section were used as the initial approximations to the phase spectrum. In Figs. 15a through 15j, the refined phase estimates and the resulting reconstructions are shown for each of the various SNR. In each case, the algorithm was stopped when the weighted correction signal energy was reduced to .16. Originally it was intended to have this be .1, but the procedure was unable to reduce the correction energy much below .16 for the two lowest SNR cases. In all cases, the exact extent of the object, or region of support, was assumed known.

Phase refinement actually degraded the results for the two poorest SNR cases. Slight improvement for the 20 dB

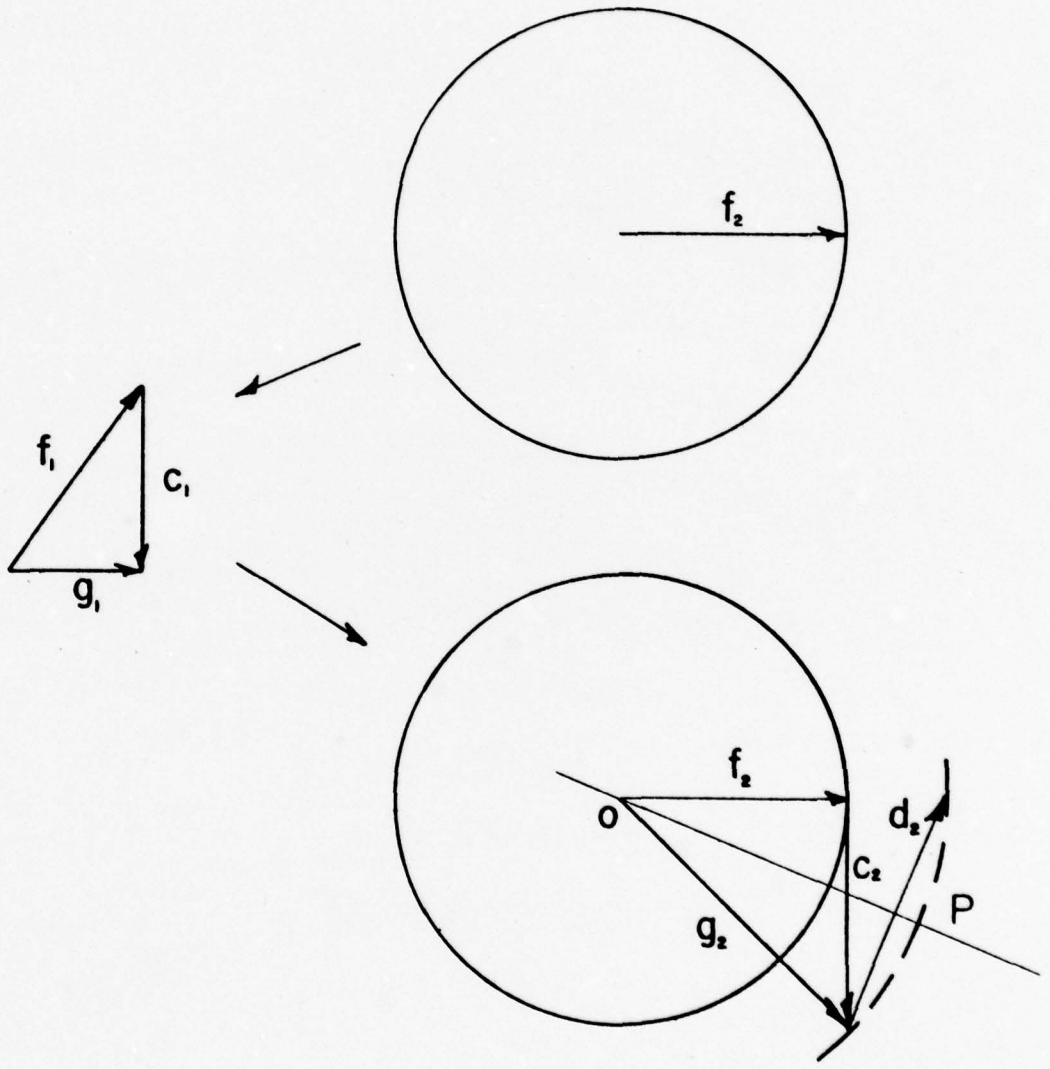
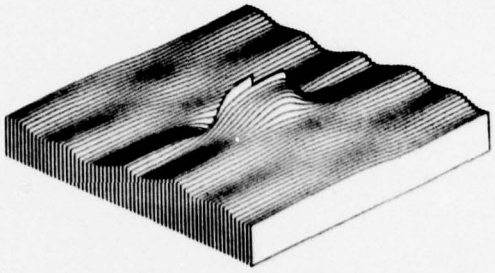


Fig. 14. A geometric proof of the nonconvergence of the Gerchberg-Saxton algorithm when used for magnitude refinement.

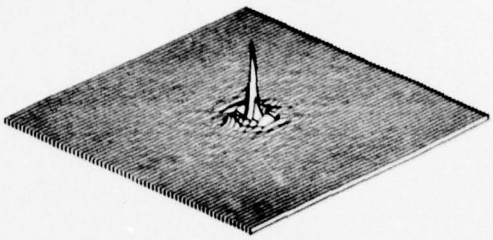
case is apparent, but for the 30 dB and 40 dB cases the improvement is substantial. The refinement required 29, 29, 7, 12 and 22 iterations respectively. Note that as the SNR gets poorer, the phase tends to be underestimated. The obvious conclusion is that the technique is not a cure-all, but it can provide real improvement if the magnitude spectrum is fairly accurate. One would expect good success if, instead of the entire magnitude spectrum, only the frequencies with locally good SNR were used to determine the phase refinement. This is correct, and it forms an important part of the super-resolution procedure that is discussed in the next chapter.

MIN=- 300745101 (- 300745101)
MAX= 314159301 (314159301)
AVGE= 7669848e-3



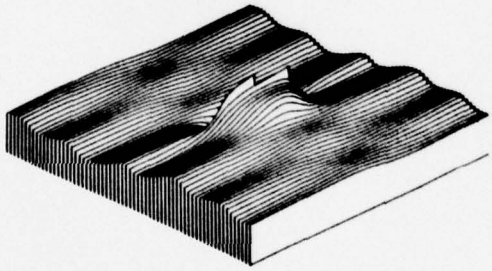
a) Phase at 5 dB SNR
WSSE = 608.187

MIN=- 1111151 (- 1111151)
MAX= 151010101 (151010101)
AVGE= 1507801e-2



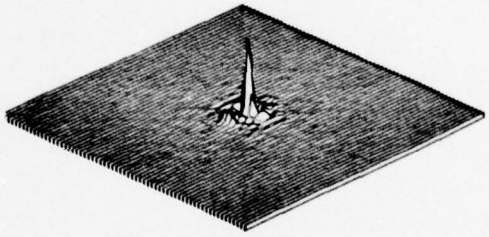
b) Object at 5 dB SNR
SSE = 1.449/1.115

MIN=- 307977801 (- 307977801)
MAX= 314159301 (314159301)
AVGE= 7669905e-3



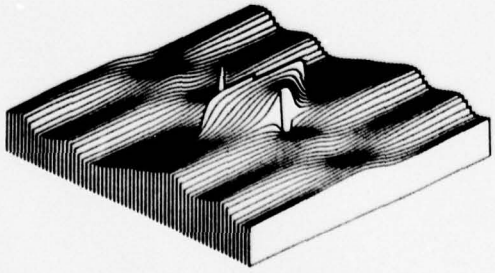
c) Phase at 10 dB SNR
WSSE = 618.252

MIN=- 1028597 (- 1028597)
MAX= 149473601 (149473601)
AVGE= 1511328e-2



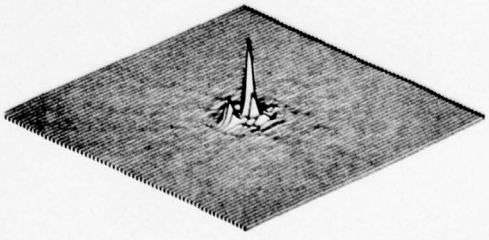
d) Object at 10 dB SNR
SSE = 1.371/1.053

MIN=- 313038001 (- 313038001)
MAX= 314159301 (314159301)
AVGE= 7669842e-3



e) Phase at 20 dB SNR
WSSE = 766.827

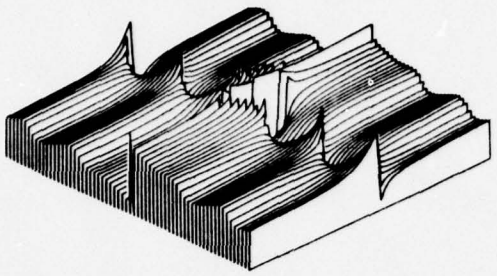
MIN=- 63663840-1 (- 63663840-1)
MAX= 127232601 (127232601)
AVGE= 1522832e-2



f) Object at 20 dB SNR
SSE = .684/.514

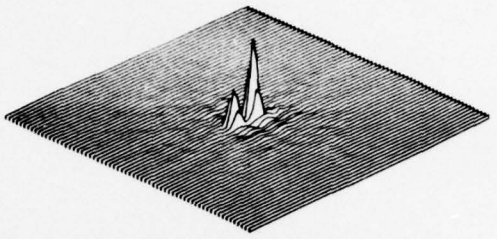
Fig. 15(a-f) Effects of Phase Refinement at Various SNR

MIN=- 3119372@1 (- 3119372@1)
MAX= 3141593@1 (3141593@1)
AVGE= 2300957@-2



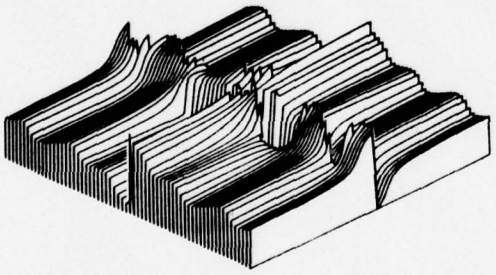
g) Phase at 30 dB SNR
WSSE = 773.167

MIN=- 4707290@-1 (- 4707290@-1)
MAX= 1015628@1 (1015628@1)
AVGE= 1526178@-2



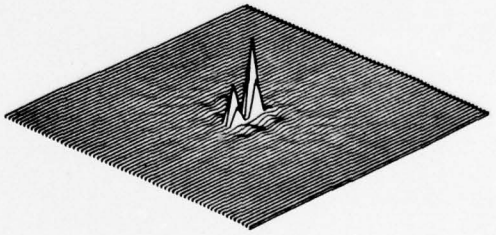
h) Object at 30 dB SNR
SSE = .269/.196

MIN=- 3139910@1 (- 3139910@1)
MAX= 3141593@1 (3141593@1)
AVGE= 2300862@-2



i) Phase at 40 dB SNR
WSSE = 414.909

MIN=- 3642376@-1 (- 3642376@-1)
MAX= 9859926 (9859926)
AVGE= 1527205@-2



j) Object at 40 dB SNR
SSE = .107/.049

Fig. 15(g-j) Effects of Phase Refinement at Various SNR

CHAPTER IV

SUPER-RESOLUTION

IV.1 Introduction

One is often confronted in a signal processing environment with a signal that has been lowpass filtered with a resulting loss of information. One would like to be able to recover this information and restore the signal. There is a strong basis in theory to support the plausibility of such recovery for a certain class of signals, those which have an analytic frequency spectrum. The difficulty in actually realizing such recovery in the practical case has led at least one pair of authors to reject the concept as a 'myth' [55]. It is my opinion that this judgement is premature and that a substantial amount of super-resolution may be had even in the case of real world blurs. This chapter discusses a new algorithm for achieving super-resolution. It is similar to the algorithm originally proposed by Gerchberg, but in our experience has demonstrated better convergence properties and much better conditioning in the presence of noise. In this chapter we will present a review of the Gerchberg algorithm, a heuristic analysis of the new algorithm, and the results of some simulations which compare the algorithms on the

synthetic blurred star data base.

IV.2 Background and Analysis of the Algorithm

The theory of analytic functions of a single complex variable is usually invoked to justify the attempts at, as well as understand the methods of, super-resolution. In an astronomical context, one is at best limited to diffraction-limited measurements. The OTF of a diffraction-limited imaging system has an absolute spatial frequency cutoff. Because a function of finite support has an everywhere analytic (or entire) z transform [42], and because an entire function must be infinite in extent, we can say that diffraction-limited imaging always removes spatial frequency information, even in the noiseless case. In reality, however, the signal energy may be sufficiently concentrated within a finite bandwidth to permit object resolution of arbitrary accuracy with a finite spectrum [44]. When this is the case, the object is said to be well resolved.

When the low-pass filtering operation of the OTF removes significant energy we say the object is poorly resolved. The classical definition of resolution specifies the minimum angular separation of two point sources which allows them to be seen as distinct, for a given optical instrument of circular pupil diameter D. This minimum separation was found to be $1.22\lambda/D$, where lambda is the wavelength of the observed light. Di Francia [51] pointed

out the practical rather than theoretical nature of this limit, however, and showed that an infinite number of object intensity distributions could produce an identical image. By this he demonstrated the implicit use of a priori knowledge on the part of the observer in determining the resolution of an optical instrument.

Harris [42] recognized that no two distinct objects of finite angular size can produce identical images. This is because a "function of a complex variable is determined throughout the entire z plane from a knowledge of its properties within an arbitrarily small region of analyticity" . Thus an entire function is everywhere determined by its behavior over an arbitrarily small region, so that an accurate measurement of a finite portion of the spectrum is sufficient to uniquely determine the complete spectrum, and hence the object, given that it has finite support. He concludes that "diffraction...imposes a resolution limit which is determined by the noise of the system, rather than some absolute criterion."

While the above reasoning no doubt gives insight into the problem, there remain some unresolved issues when one is constrained to computer solutions. These seem to me to include the following questions. First, is the requirement that complete knowledge of the function over some region of analyticity satisfied by knowledge of the Fourier transform of the function, which is at best knowledge along a line or

contour in the complex domain? Second, is the requirement satisfied when one has only discrete samples of the Fourier transform? Third, does the proof of the Paley-Weiner theorem extend to functions of more than a single complex variable? In spite of these questions, the semi-quantitative analysis given here will depend on the properties of analytic functions. An analysis along the lines of that given by Youla [48] would be independent of the questions just raised, however, and would be highly desirable.

Since our new algorithm is an adaptation of the one due to Gerchberg, a review of the latter will help in understanding our work and in placing it in proper perspective. The Gerchberg algorithm iterates between the spatial and frequency domains. In the spatial domain, the reconstruction is set to zero outside the known region of support of the object, and in the frequency domain the newly estimated spectrum is always corrected to agree with the known portion of the spectrum. A flow chart of the algorithm is found in Fig. 16. The algorithm is shown to converge uniquely for noiseless data. Gerchberg recommends a procedure to limit the buildup of noise in the reconstruction due to noise in the 'known', or measured, portion of the spectrum. This operation is necessary because correcting the known portion of the spectrum adds back all the noise energy originally present, so that the

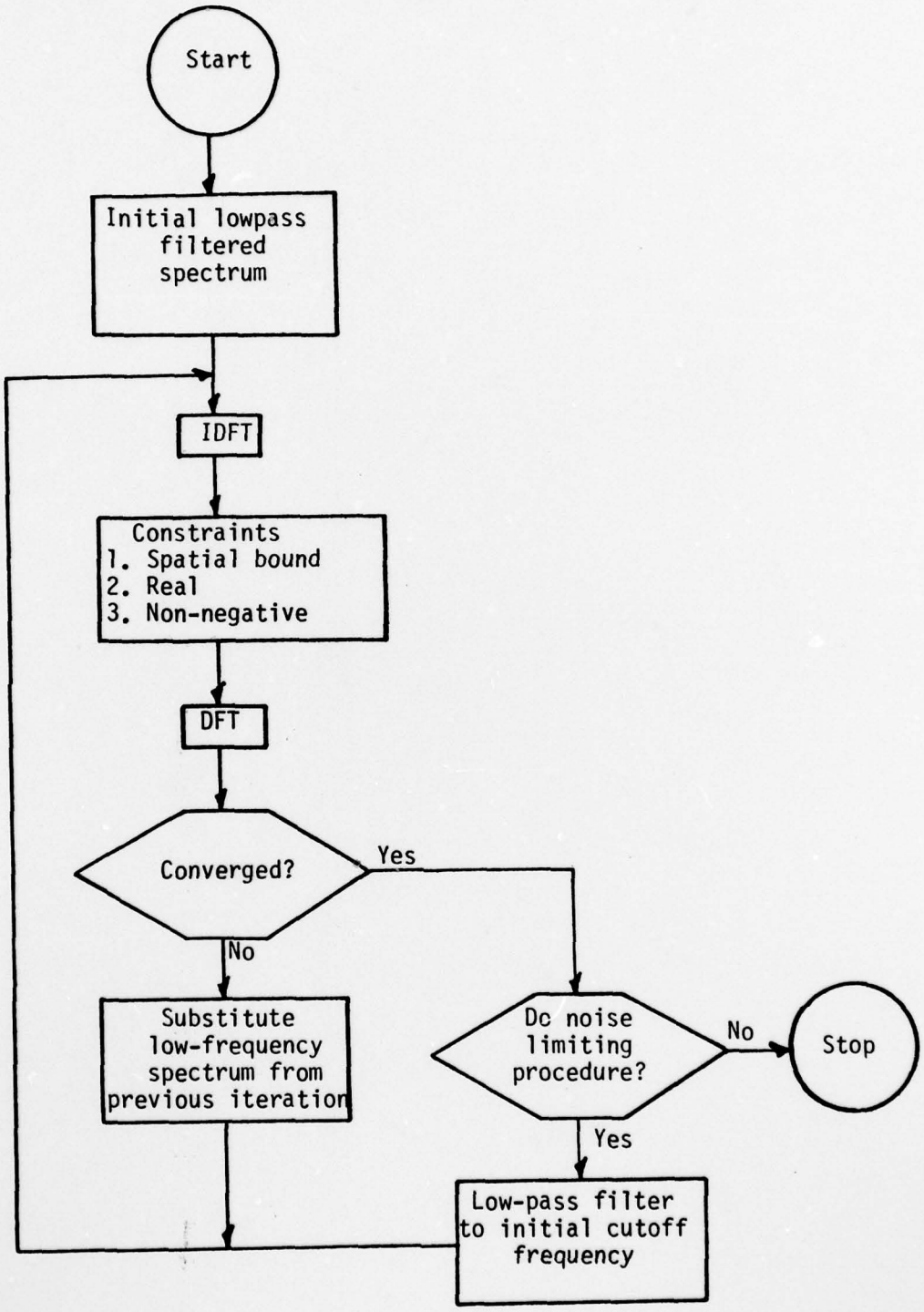


Fig. 16. Flow chart of Gerchberg super-resolution algorithm, including the noise-limiting procedure.

error energy due to this noise must increase or remain constant at every iteration. At some point the error due to the increase in this noise energy becomes greater in magnitude than the decrease in error energy resulting from refining the extrapolated spectrum. At this point the algorithm should be terminated. This result has been shown [48] on theoretical grounds for the continuous case. Unfortunately, the optimal number of iterations can not be known a priori.

The noise limiting procedure is based on the observation that energy in the converged reconstruction which lies outside the known object extent must be due to the original distortion energy. Therefore, the object is limited to its correct spatial extent, the transform taken, and the new lowpass spectrum is used to start the iteration all over again. The procedure thus eliminates that portion of the original spectral noise energy which lies outside the known object extent.

The new algorithm corrects this deficiency by not correcting the magnitude spectrum to the measured magnitude spectrum, so that measurement error is not increased. Instead the magnitude and phase spectra are alternately extended through the frequency space. The phase spectrum extension is allowed at each step to converge in a weighted SSE sense, where the weighting function is the estimated magnitude spectrum. A flowchart is illustrated in Fig. 17.

As shown in Chap. III.6, no magnitude iterations are performed since this is an unstable procedure. As each frequency extension is made the total energy in the spectrum below the old frequency cutoff changes. If the new magnitude spectrum is simply scaled to agree with the old spectrum (in whatever sense is appropriate) below the old cutoff frequency, the effective convergence of the algorithm is easily determined by the amplitude of this scaling constant. In my work, the algorithm was always terminated when something in excess of 99% of the total spectral energy was estimated. The phase estimates exhibited erratic convergence behavior when frequency extrapolation was attempted much beyond that point. Although their convergence could often be stabilized by a smaller cutoff frequency extension, the computational time simply became excessive.

There seem to be two sources of error in the extrapolated spectrum produced by the new algorithm. They are induced by the way in which the algorithm treats the two kinds of errors in the original measured spectrum. Following Gerchberg, we call these two kinds error energy and distortion energy.

We can consider the initial lowpass spectrum estimate to be the sum of the correct spectrum and an error spectrum which exists only beyond the initial lowpass cutoff frequency. The error spectrum is equal in magnitude and

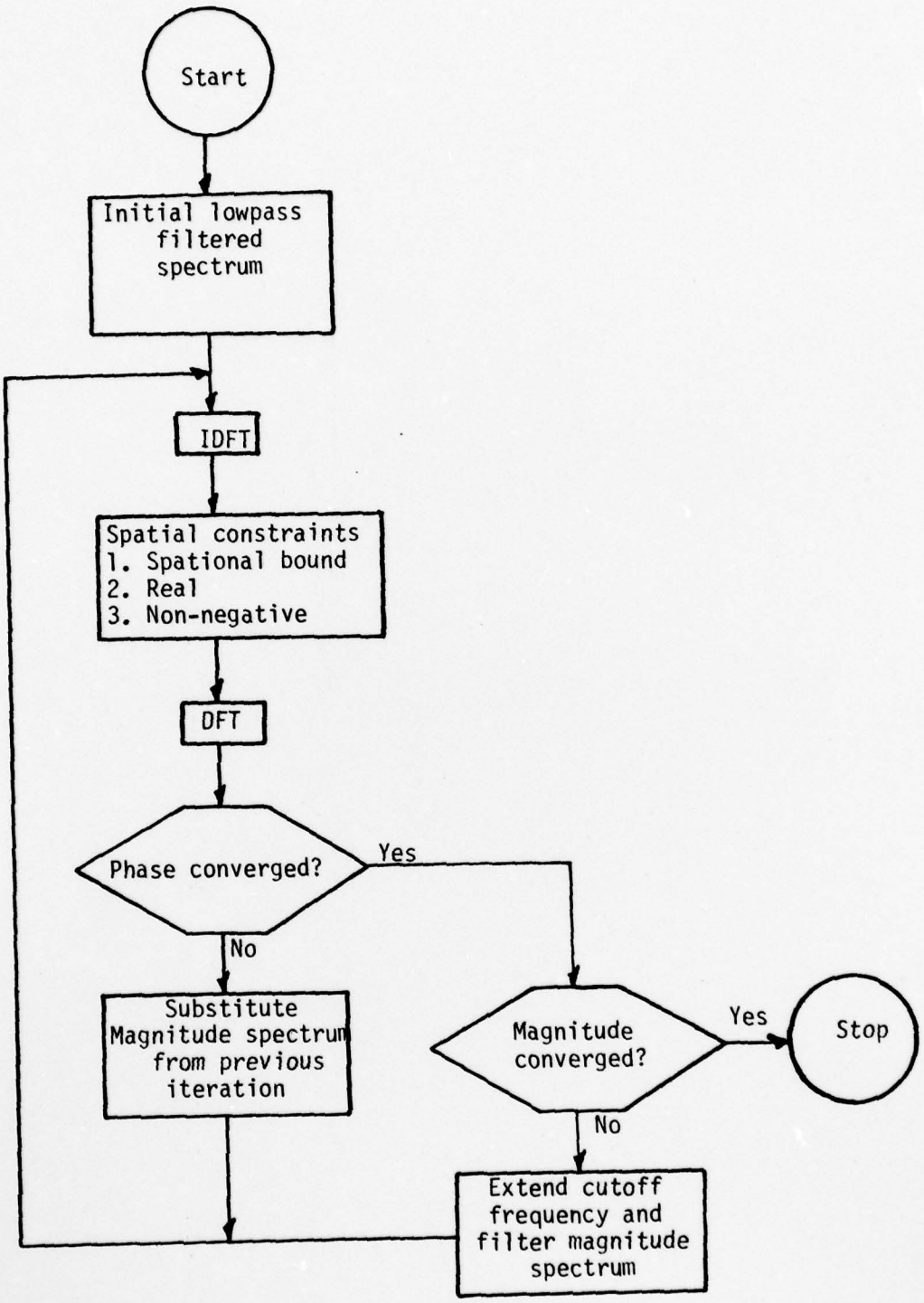


Fig. 17. Flow chart of the new nonlinear super-resolution algorithm.

opposite in sign to the true spectrum. Inside the cutoff frequency, the known spectrum is represented as the sum of the true spectrum and a distortion spectrum. Consider first the case where the distortion energy is zero. Since the error spectrum is band-limited, it follows that its inverse Fourier transform, the error object, cannot be spatially limited. Thus some error energy will be preserved when the spatial constraints are imposed. Consequently the spectrum of the constrained object will retain some portion of the original error energy, some of which may be found even in the low frequencies below the initial cutoff frequency. Since the low frequencies of this new spectral estimate are retained in the magnitude spectrum extension, the algorithm has introduced into the low frequency spectrum a portion of what was originally error energy. To minimize the growth of this error we should minimize the number of times we extend the spectrum. This implies either rapid spectral extrapolation or that we confine ourselves to signals that are nearly resolved so that few extrapolations are required. Because most images have a general lowpass characteristic, one would expect the error energy injected into the low frequencies to decrease as the cutoff frequency is increased and the remaining error energy is decreased.

Suppose now that distortion energy is present. It is probable that the spatial constraints will also preserve

some portion of this energy, as well. The spectrum of the newly-constrained object may similarly have distortion energy present in all frequencies. To the extent that distortion energy is injected into the high frequencies, Parseval's theorem indicates that the total distortion energy must be decreased by lowpass filtering, even with a higher cutoff frequency. To minimize the distortion energy in the final result we want to perform many filtering operations, which implies slow spectral extrapolation. Changing the rate of extrapolation thus varies the relative contribution of the distortion and error energies to the final extrapolated estimate. Although it is possible that a variable rate would improve the algorithm performance, all the experimentation performed here had a fixed rate of frequency cutoff increase. The optimum rate of increase or even its existence in the general case, is not presently known.

The heuristic analysis just presented is linear, and in the sense that at every step the algorithm is dealing with disjoint regions in the frequency domain, it is a linear algorithm. However the possible importance of the nonlinearity introduced by the separate extension of the magnitude and phase is not yet known. In summary, the original Gerchberg algorithm is able in the limit to decrease the error energy to zero at the expense of increasing the total distortion energy. The new algorithm

sacrifices the possibility of perfect reconstructions of noise-free data in exchange for stable behavior of the distortion energy.

IV.3 Super-resolution of Simulated Data

I have examined by simulation the behavior of the algorithm as the rate of increase of the cutoff frequency is changed, although the examination was certainly not exhaustive. The results on the 5 dB SNR case are summarized in Table III, where it appears that the optimum increase is about 5 discrete frequencies at each step, at least for this data. This same rate of increase was used in other simulations where the effects of SNR and initial lowpass cutoff frequency were investigated. In all the simulations, the lowpass filter is a radially symmetric Fourier lowpass filter. The phase convergence criterion was that

$$\varepsilon > \sum_{\bar{u}} |\hat{I}_o(\bar{u})| \{ \hat{\phi}^{(k)}(\bar{u}) - \hat{\phi}^{(k-1)}(\bar{u}) \}^2$$

Since our heuristic analysis indicates that some error must always remain in the solution, and may in fact increase in the 'known' portion of the spectrum, it was necessary to observe the seriousness of this characteristic in a practical setting. A number of reconstructions based on lowpass filtering the non-noisy data at various cutoff frequencies were effected. They are shown together with the initial estimates and their super-resolved spectra in

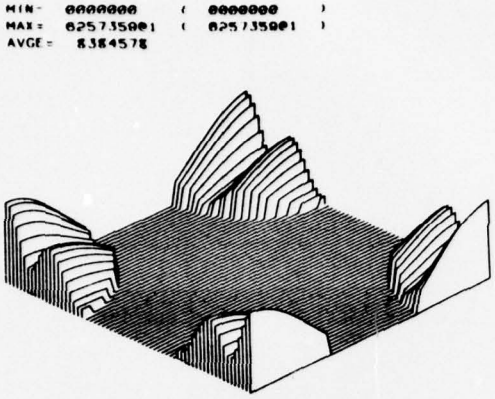
Table 3. Variation of super-resolved image SSE in the 5-dB SNR reconstruction with different initial frequency cutoffs and frequency extrapolation step sizes.

Initial Frequency Cutoff (Cycles/Picture)	Step Size (Cycles/Picture)	Number of Iterations	Reconstructed Image SSE
10	20	16	.484
10	5	31	.251
10	3	55	.320
5	5	26	.308
5	10	17	.326
8	5	37	.016
10	10	30	.330

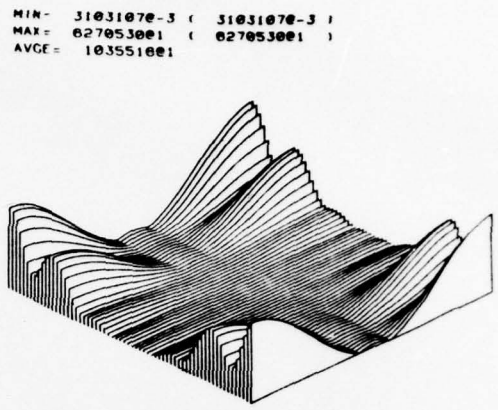
Figs. 18 a) through 18 t). In every case, $\varepsilon = .1$. The reconstructions are quite stable with decreasing initial frequency cutoff, although it is curious to note that the last estimate has a slightly better SSE than the third estimate. The reconstructions required 10, 15, 18, and 22 iterations, respectively. All of the reconstructions have errors less than 2.5%. It appears that the error introduced into the 'known' portion of the spectrum is relatively small and well behaved.

Figs. 19 a) through 19 y) show the super-resolution of the least-squares phase Knox-Thompson reconstructions of chapter III. It was necessary to use a lower and lower initial cutoff frequency as the SNR got poorer, so as to maintain a reasonably accurate initial estimate. Note that even in the 5 dB SNR case an excellent reconstruction is produced. In most cases the peak magnitudes, while about correct relative to each other, are in error absolutely. The information in the peaks resides almost entirely in the highest frequencies, where less than one percent of the signal energy is concentrated. Misestimation in these frequencies is thus easily detected in the spatial domain. (The final estimates have been scaled to have total energy equal to that of the true object). None of the solutions required more than 33 iterations.

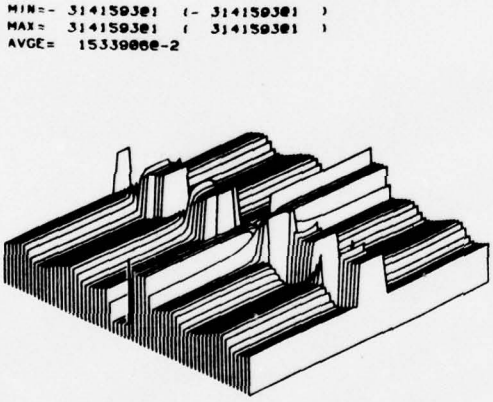
The performance of the procedure seems to depend not only on the amount of noise in the initial estimate but



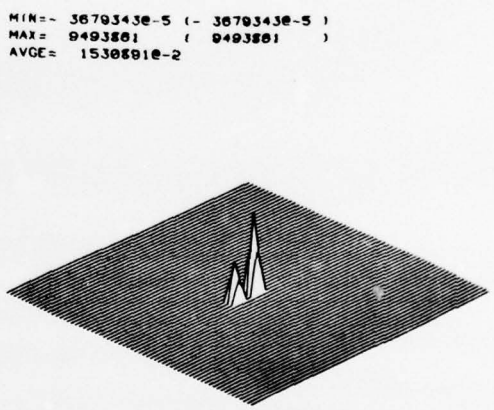
a) Initial Magnitude
94.6%



b) Extrapolated Magnitude
SSE = 25.331

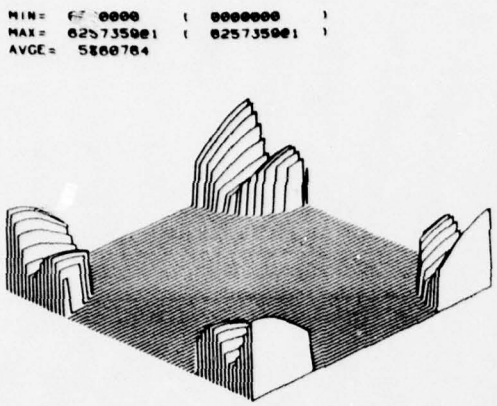


c) Extrapolated Phase
WSSE = 253.530

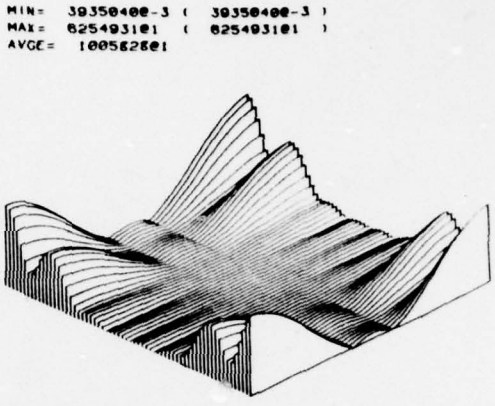


d) Super-resolved Object
SSE = .0075

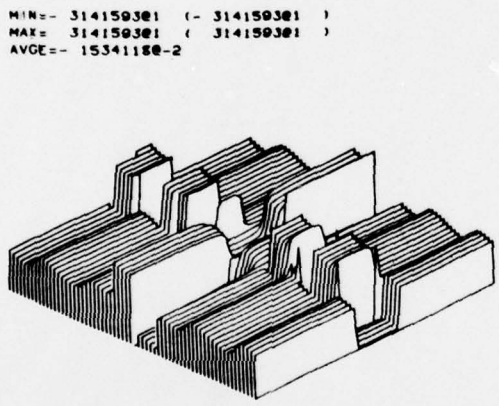
Fig. 18(a-d) Super-Resolution of Noise-Free Data with Initial Cutoff Frequency at 20 Cycles/Picture



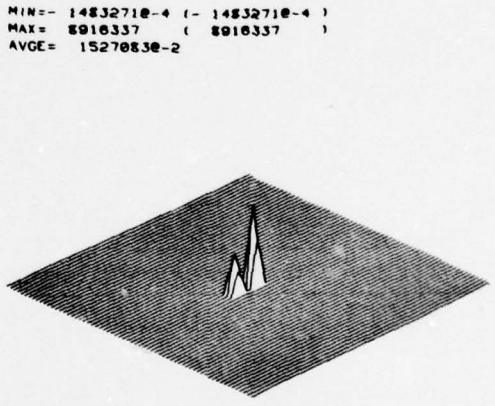
e) Initial Magnitude
77.3%



f) Extrapolated Magnitude
SSE = 60.826



g) Extrapolated Phase
WSSE = 317.232



h) Super-resolved Object
SSE = .018

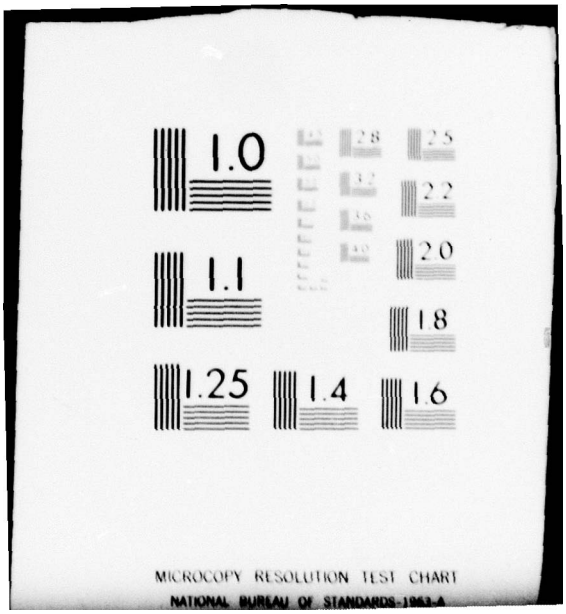
Fig. 18(e-h) Super-Resolution of Noise-Free Data with Initial Cutoff Frequency at 15 Cycles/Picture

AD-A073 434 UTAH UNIV SALT LAKE CITY F/G 3/1
HIGH RESOLUTION ASTRONOMICAL IMAGING THROUGH THE TURBULENT ATMO--ETC(U)
JUL 79 R L FROST, C K RUSHFORTH, B S BAXTER AFOSR-77-3212
UNCLASSIFIED UTEC-79-081 NL

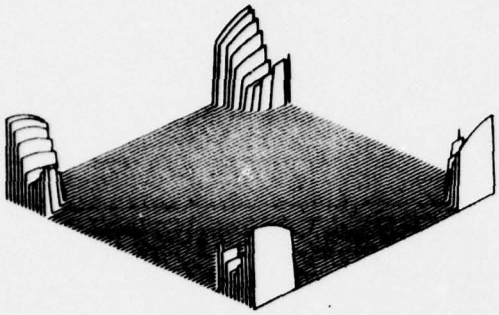
2 OF 2
AD A073434



END
DATE FILMED
10-79
DDC

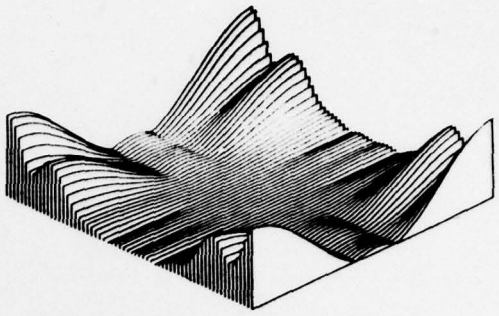

1.0
1.1
1.25
1.4
1.6
2.0
2.2
2.5
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MIN= 0000000 (0000000)
MAX= 025735901 (025735901)
AVGE= 2032240



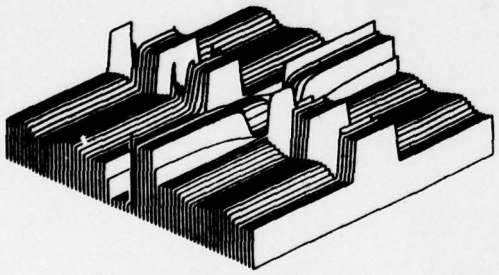
i) Initial Magnitude
43.2%

MIN= 27906820-4 (27906820-4)
MAX= 025335901 (025335901)
AVGE= 1089514@1



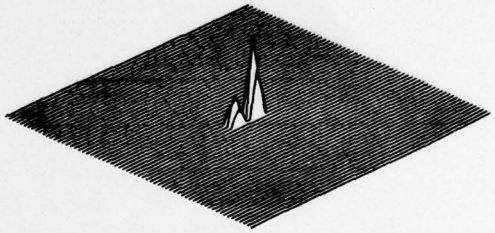
j) Extrapolated Magnitude
SSE = 146.609

MIN=- 314159301 (- 314159301)
MAX= 314159301 (314159301)
AVGE= 15339840-2



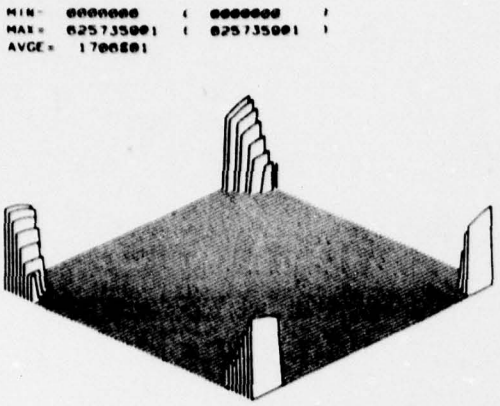
k) Extrapolated Phase
WSSE = 402.060

MIN=- 10583150-4 (- 10583150-4)
MAX= 10004700! (10004700!)
AVGE= 1526707@-2

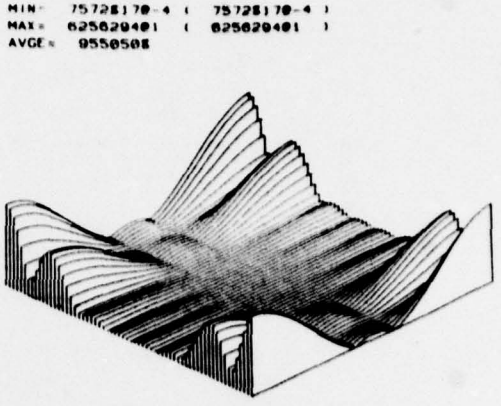


l) Super-resolved Object
SSE = .069

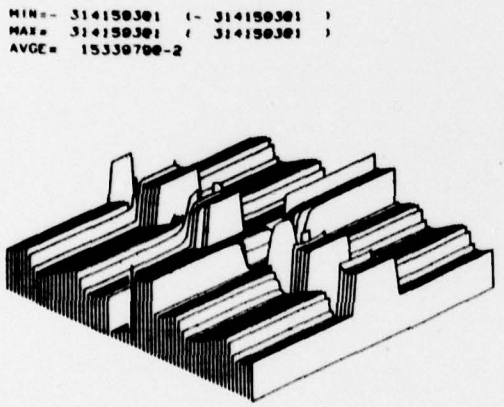
Fig. 18(i-l) Super-Resolution of Noise-Free Data with Initial Cutoff Frequency at 10 Cycles/Picture



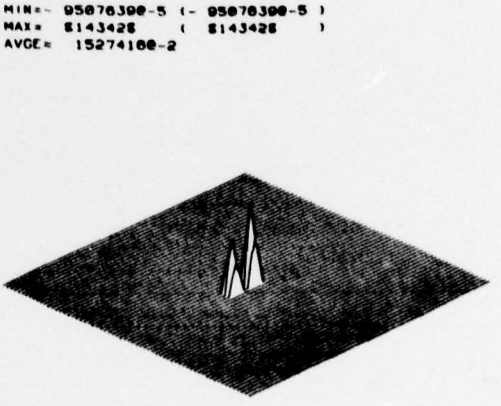
m) Initial Magnitude
28.7%



n) Extrapolated Magnitude
SSE = 135.327

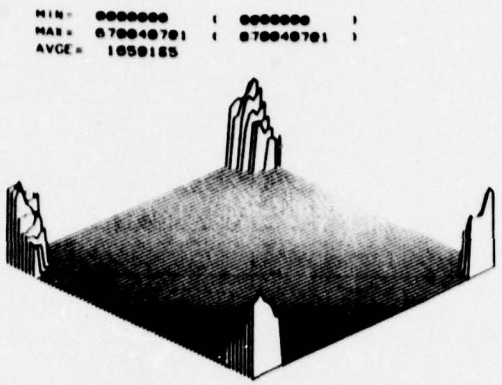


o) Extrapolated Phase
WSSE = 361.243

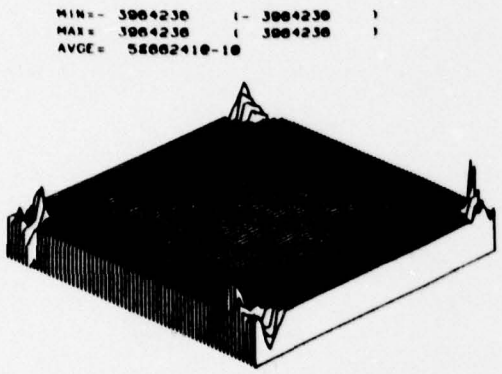


p) Super-resolved Object
SSE = .061

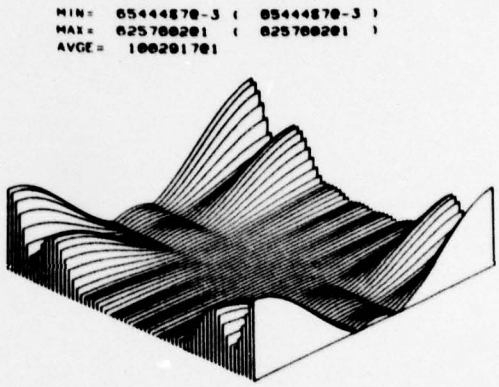
Fig. 18(m-p) Super-Resolution of Noise-Free Data with Initial Cutoff Frequency at 7 cycles/Picture



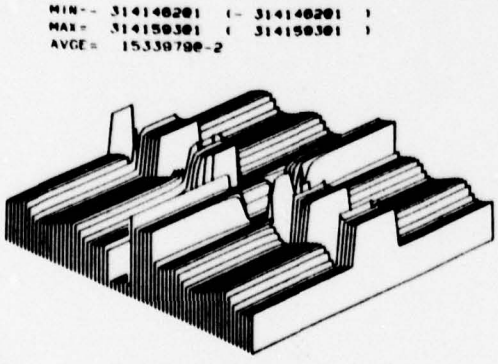
a) Initial Magnitude



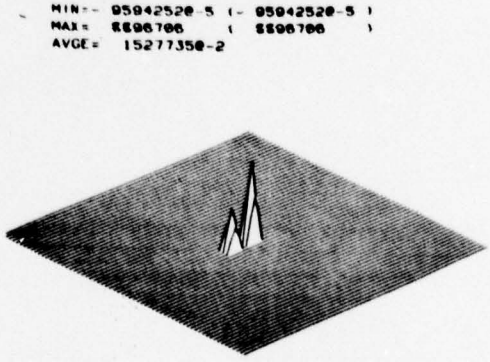
b) Initial Phase



c) Extrapolated Magnitude
SSE = 60.901

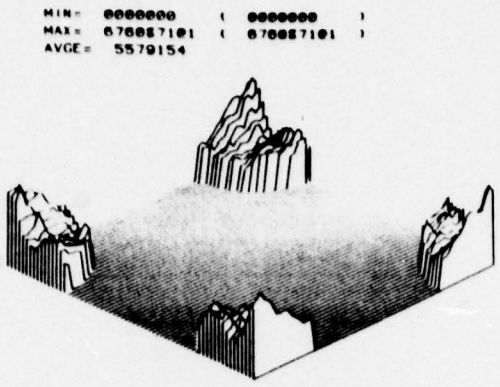


d) Extrapolated Phase
WSSE = 302.583

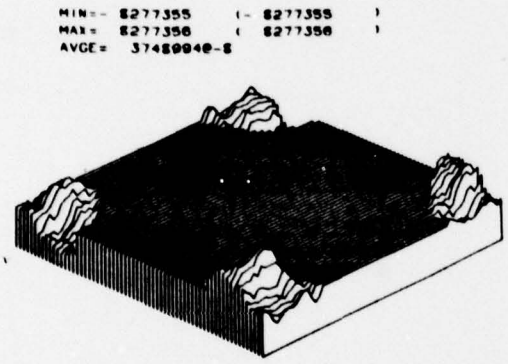


e) Super-resolved Object
SSE = .018

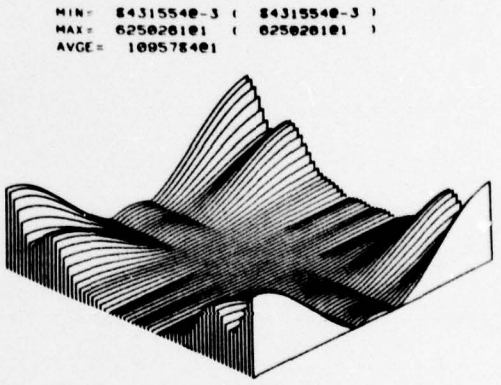
Fig. 19(a-e) Super-Resolution of Noisy Data at 5 dB SNR with Initial Cutoff Frequency at 7 Cycles/Picture



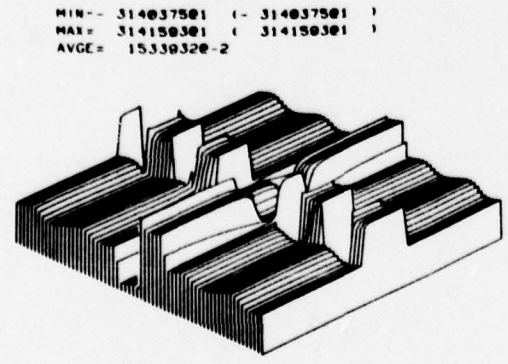
f) Initial Magnitude



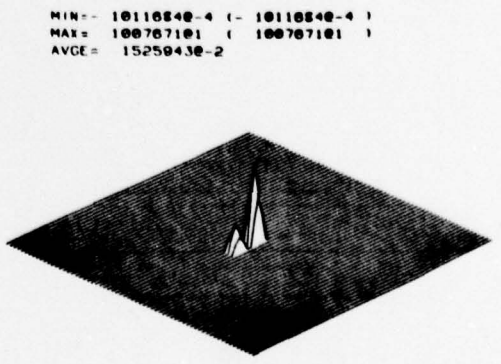
g) Initial Phase



h) Extrapolated Magnitude
SSE = 160.181

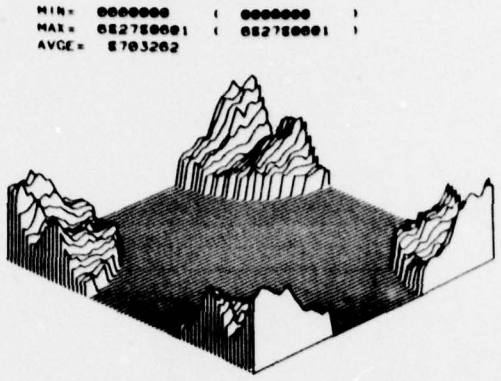


i) Extrapolated Phase
WSSE = 394.871

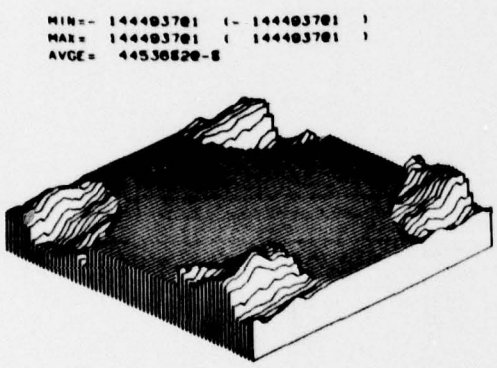


j) Super-resolved Object
SSE = .076

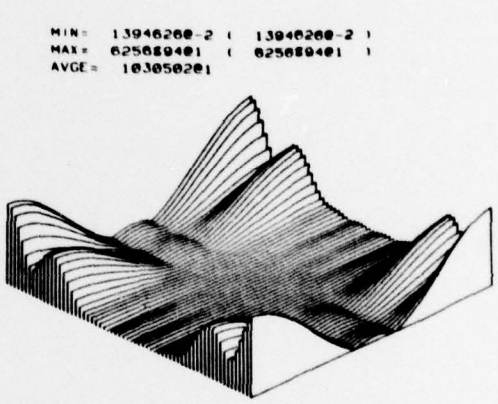
Fig. 19(f-j) Super-Resolution of Noisy Data at 10 dB SNR with Initial Cutoff Frequency at 15 Cycles/Picture



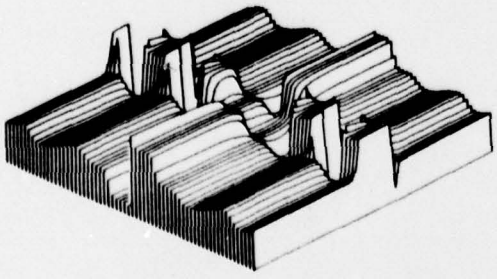
k) Initial Magnitude



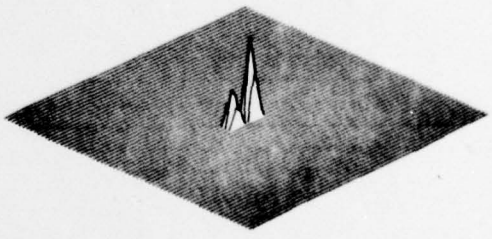
l) Initial Phase



m) Extrapolated Magnitude
SSE = 88.321

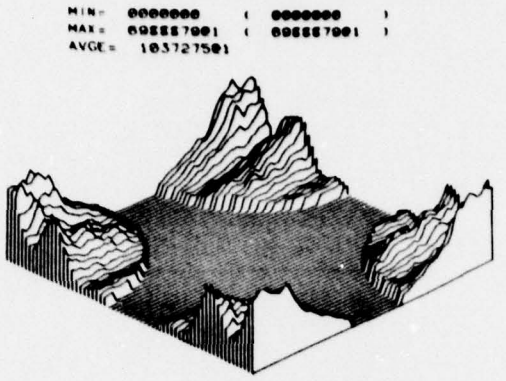


n) Extrapolated Phase
WSSE = 273.525

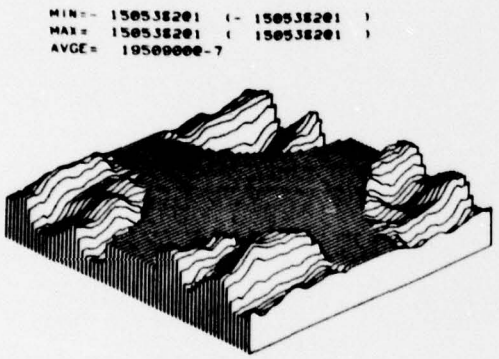


o) Super-resolved Object
SSE = .032

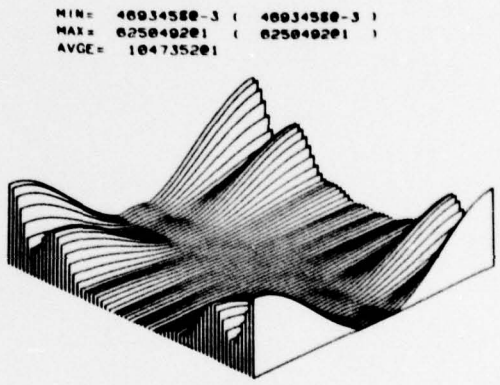
Fig. 19(k-o) Super-Resolution of Noisy Data at 20 dB SNR with Initial Cutoff Frequency at 20 Cycles/Picture



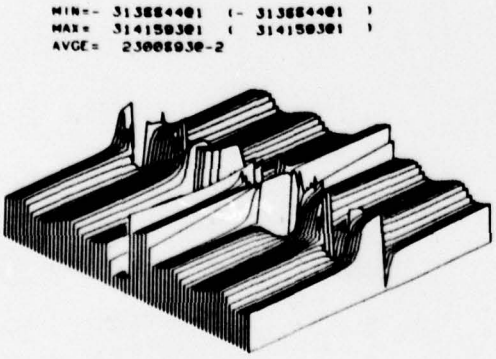
p) Initial Magnitude



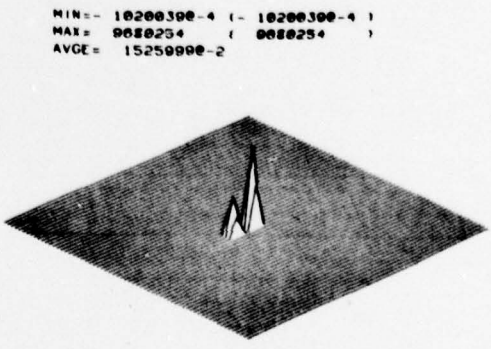
q) Initial Phase



r) Extrapolated Magnitude
SSE = 29.044

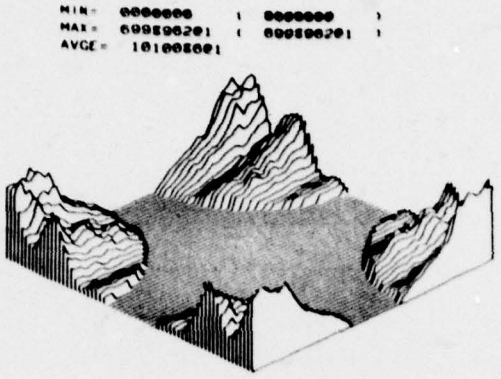


s) Extrapolated Phase
WSSE = 280.967

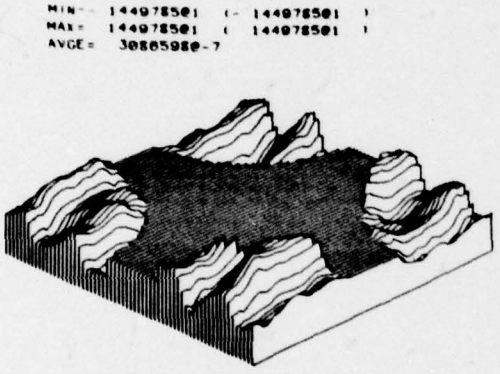


t) Super-resolved Object
SSE = .013

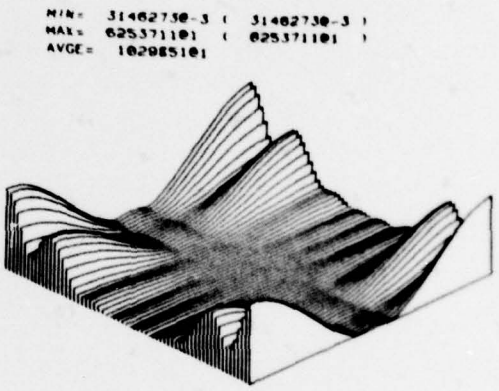
Fig. 19(p-t) Super-Resolution of Noisy Data at 30 dB SNR with Initial Cutoff Frequency at 25 Cycles/Picture



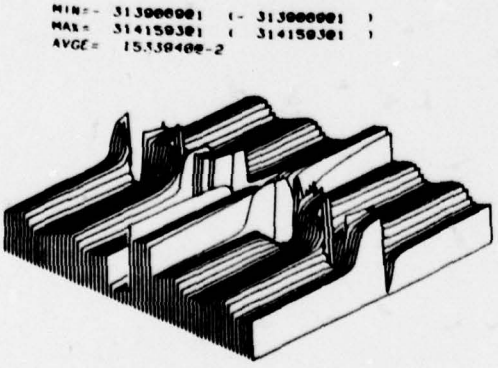
u) Initial Magnitude



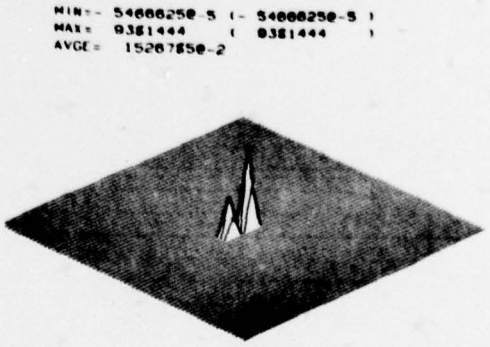
v) Initial Phase



w) Extrapolated Magnitude
SSE = 34.491



x) Extrapolated Phase
WSSE = 278.857

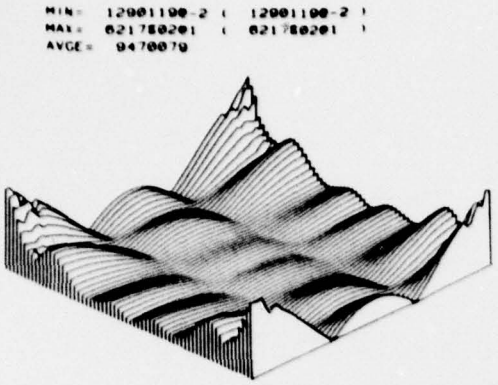


y) Super-resolved Object
SSE = .013

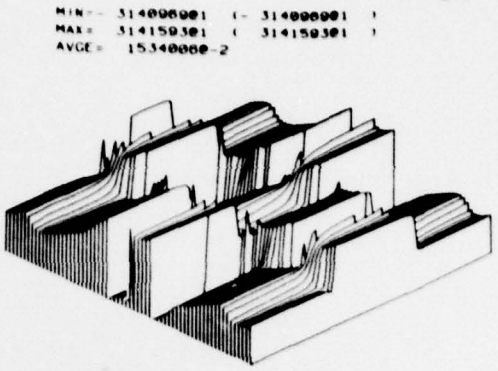
Fig. 19(u-y) Super-Resolution of Noisy Data at 40 dB SNR with Initial Cutoff Frequency at 25 Cycles/Picture

also on its structure, so that an apparent increase in local SNR may not necessarily imply a more accurate reconstruction. The spectral estimates are smoothed by the procedure, as shown in the illustrations. Particularly striking is the fact that the distortion energy in the initial magnitude spectrum has been decreased in the corresponding lowpass portion of the extrapolated spectrum to between 1% and 11% of its original value. However, in extrapolating real frequency data, situations might arise where one would prefer to replace the final low frequency spectrum with the measured spectrum to avoid as much as possible this smoothing.

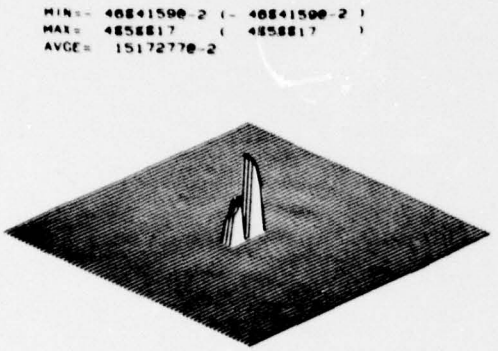
As an illustration of the superiority of the new algorithm a reconstruction of the 5 dB case using Gerchberg's algorithm is included. The same initial lowpass filter was used, and the phase was refined, so that for both algorithms the 'known' portion of the spectrum was identical. In Figs. 20 a-c) a solution obtained by allowing 150 iterations, with the noise limiting procedure on the 101st iteration, is shown. This reflects the common usage of the algorithm. In Figs. 20 d-f) the optimal number of iterations was allowed, that is, 29 iterations with the noise limiting procedure on the 15th iteration. (It is clear that this determination of the optimum number was possible only because the answer was already known). In both cases very inferior reconstructions were produced.



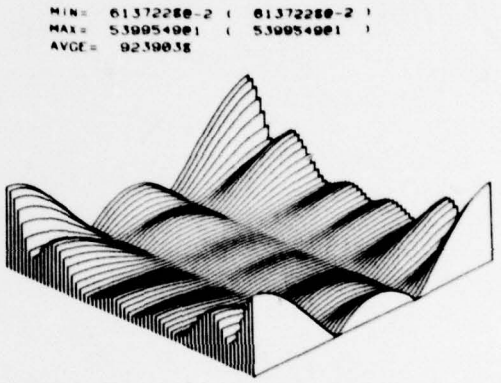
a) Extrapolated Magnitude
SSE = 1438.376



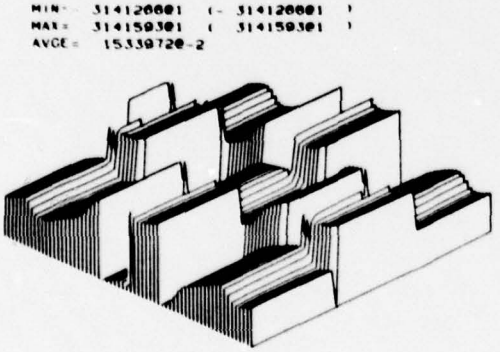
b) Extrapolated Phase
WSSE = 2454.821



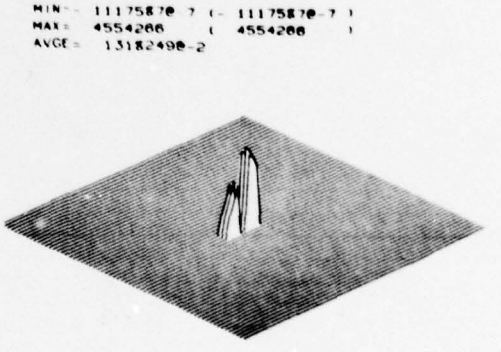
c) Super-resolved Object
SSE = .580



d) Extrapolated Magnitude
SSE = 1369.86



e) Extrapolated Phase
WSSE = 775.785

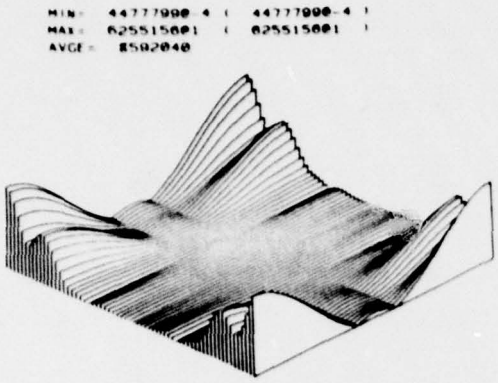


f) Super-resolved Object
SSE = .537

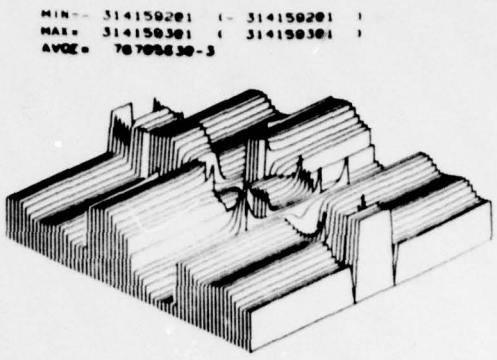
Fig. 20 Super-Resolution of Noisy Data at 5 dB SNR by Gerchberg's Algorithm

This disparity lessens as the quality of the initial data improves, but the superior convergence properties of the new algorithm give it the advantage in these cases as well. As the noise in the data decreases, the amplification of the noise by Gerchberg's algorithm does not become a serious problem until the error spectrum energy is reduced to a lower level. Thus the optimal number of iterations increases. It is not uncommon to use several hundred iterations to approximate convergence. The new algorithm seems always to converge in about the same number of iterations for a given initial cutoff frequency, nearly independently of SNR.

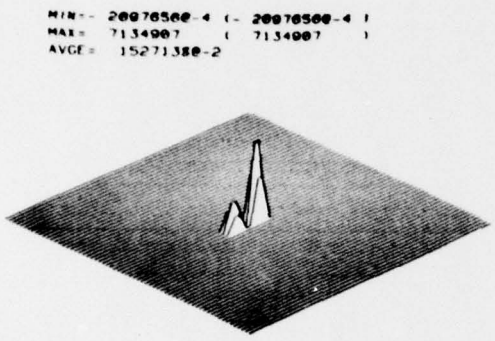
To test the sensitivity of the reconstructions to inaccuracies in determining the region of support a number of mask sizes was used. The true region of support was 18 pixels. Illustrated in Figs. 21 a) through 21 f) are reconstructions based on initial non-noisy data using masks of area 24 and 32 pixels, respectively, assuming an initial cutoff frequency of 7 cycles per picture, as in Figs. 18 p-q). The resulting reconstructions have substantial error, although the binary star character of the object is still clearly visible. Of course, the changes in the lowpass magnitude spectrum are indicative of the error in estimating the region of support. We have also observed that the phase convergence can be adversely affected if the region of support is poorly known. The



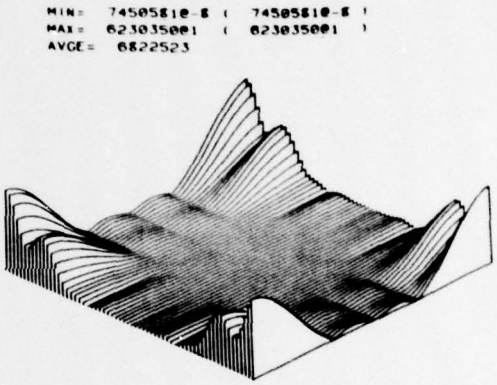
a) Extrapolated Magnitude
SSE = 1024.791



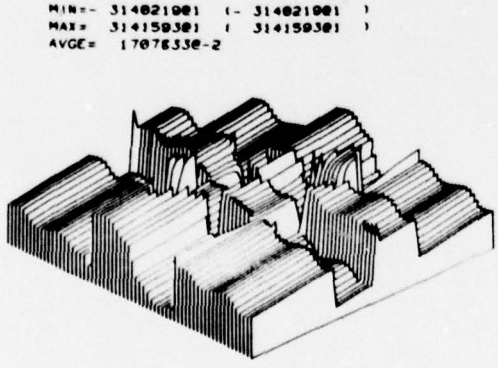
b) Extrapolated Phase
WSSE = 1776.414



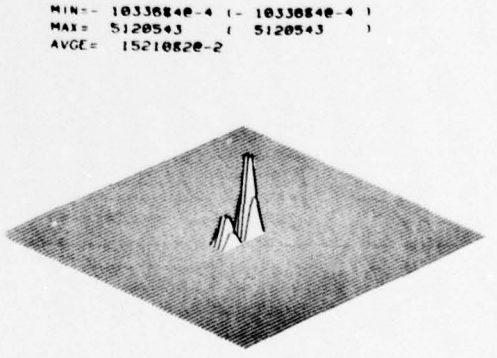
c) Super-resolved Object
SSE = .534/.306



d) Extrapolated Magnitude
SSE = 1744.553



e) Extrapolated Phase
WSSE = 2813.536



f)Super-resolved Object
SSE = .946/.612

Fig. 21 Super-Resolution of Noise-Free Data Using Inaccurate Spatial Constraints

importance of an accurate estimate of the region of support is certainly indicated, although the degradation with increasing mask size is graceful.

In super-resolving real data, it might be difficult to accurately estimate the region of support based on only the estimated autocorrelation of the object, which will itself have been low-pass filtered. It is quite possible that a priori knowledge may be needed to determine the object size and shape. An alternative would be simply to iteratively approximate the region of support until one determined which region minimized the change in the 'known' portion of the magnitude spectrum, subject to whatever knowledge is available concerning the object itself.

CHAPTER V

RESTORING TURBULENCE-BLURRED STARS

V.1 Description of Data Base

The techniques described in the preceding chapters have been applied to turbulence-blurred images of the stars α-ORI (Betelgeuse) and α-AUR (Capella). The point star reference used to estimate to mean-square OTF was β-ORI. Betelgeuse is a red super-giant star with the largest apparent angular diameter of any single star in the heavens. Capella is a binary star whose maximum interstar distance is less than .06 arc-seconds, which is comparable to diameter of Betelgeuse. Capella is particularly interesting in that it is unresolved in our data.

Our data base consists of 180 images of each star. The observations were made with the 160-inch Mayall telescope at the Kitt Peak Observatory in Arizona, by S.P. Worden and B.S. Baxter, on the nights of June 13 and 14, 1976. The optical interference filters had a central wavelength of 6500 Å and a bandpass of 20 Å. Exposure times were about 50 ms, with approximately one second between exposures to allow the atmospheric point spread functions to decorrelate temporally. The data was originally recorded on film, grey scale corrected,

processed on an image intensifier, sampled and the image intensity was digitized to 12 bits on a 256*256 grid. The scanning aperture was square with width 100 microns and the sampling interval was 50 microns. The resulting scale is .01 arc-seconds/grid interval, for a square field of view 2.56 arc-seconds on a side. Since at this wavelength the absolute frequency cutoff of the telescope is about 27 cycles/arcsecond, or about 69 cycles/picture, the speckle images are oversampled by a factor of about 1.85.

V.2 Betelgeuse reconstruction

A typical speckle image, this one of Betelgeuse, is shown in Fig. 22. The corresponding Labeyrie magnitude estimate is shown in Fig. 23, while the least-squares phase estimate derived from processing 45 unwindowed speckle images is shown in Fig. 24. The corresponding spatial image is in Fig. 25. (This latter has been interpolated by surrounding the spectrum with data points equal to zero. The effect is to increase the resolution of the DFT, as is well known. In this case, the zero-padded array was 16 times larger in area, increasing the DFT resolution by a factor of 4.) The failure of the technique is evident.

The reconstruction of Figs. 26 a-c) is based on the same speckle images, this time using a spatial window. The window was a rotated version of a Hamming window of radius 40 points. The bulk of the degradation is now removed, although the object is insufficiently resolved. Much of



Fig. 22 Speckle Photograph of Betelgeuse

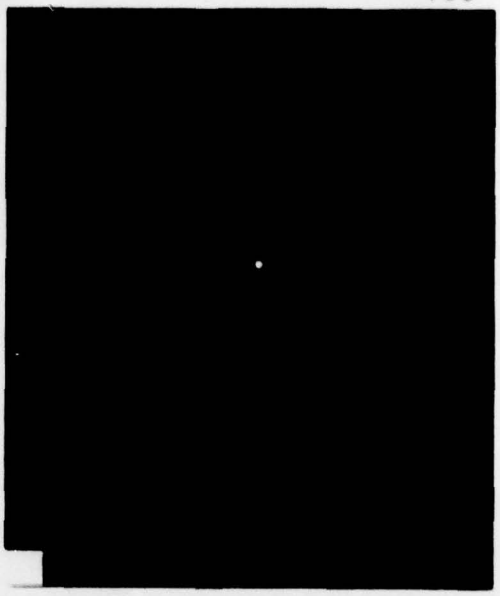


Fig. 23 Labeyrie Estimate of Object Autocorrelation

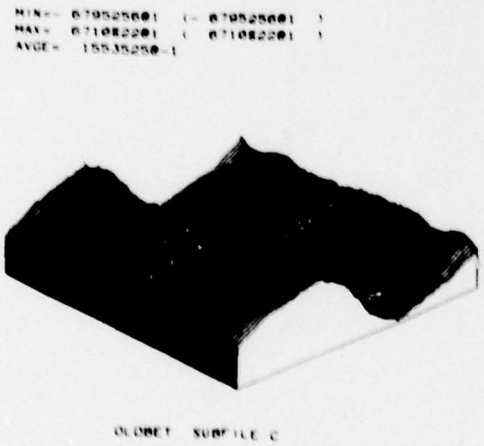


Fig. 24 Phase Estimated from Unwindowed Photographs

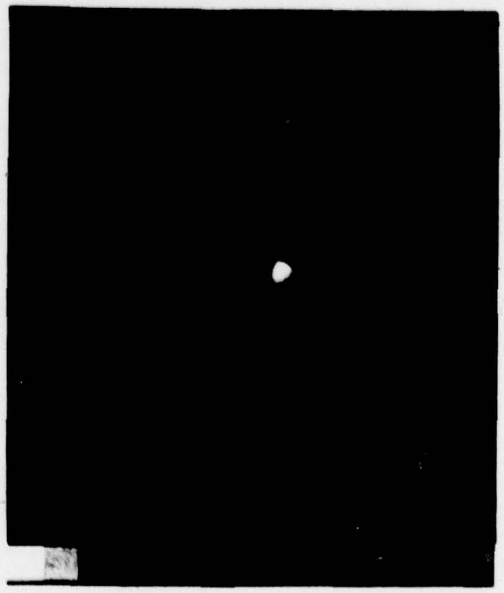
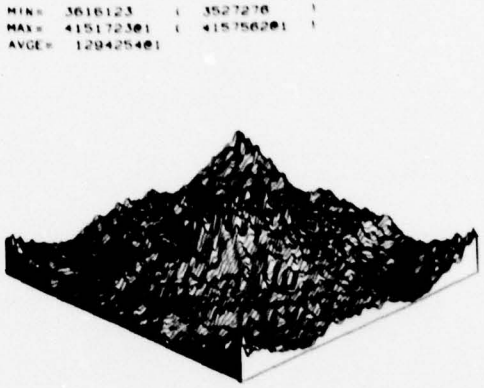
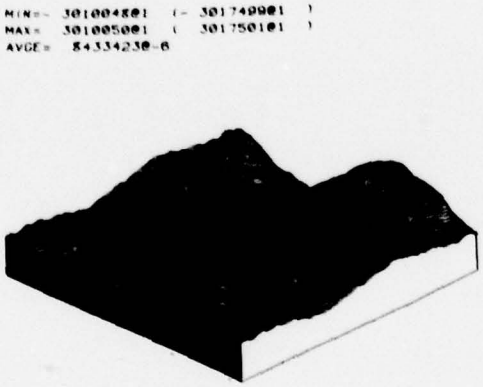


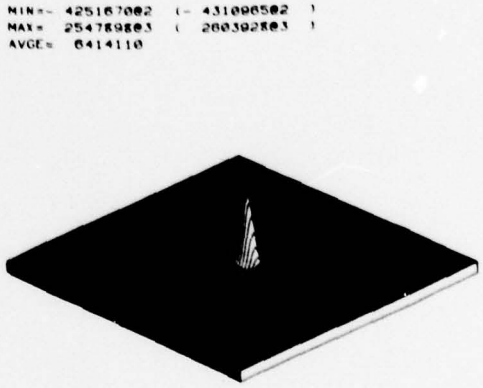
Fig. 25 Magnified Betelgeuse Reconstruction using Phase Estimate Shown in Fig. 24



a) Labeyrie Magnitude Spectrum Estimate



b) Least-Squares Phase Estimate



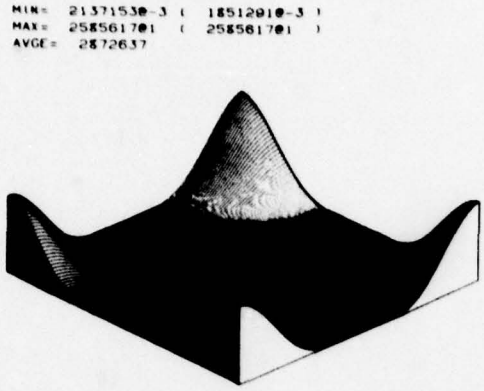
c) Estimated Intensity Object

Fig. 26 Betelgeuse Reconstructions Obtained from Hamming Windowed Speckle Photographs

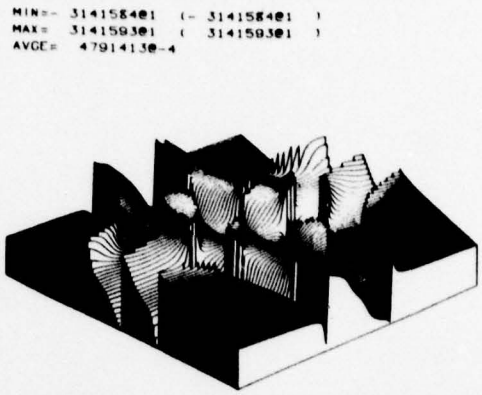
the problem is due to the rather poor magnitude characteristics of the final digital lowpass filter which is applied to eliminate the noise present in frequencies above the diffraction limit. To minimize the annoying artifacts of ringing induced in the image by the filtering operation, one is constrained to filters with gentle roll-off characteristics. Hence, additional smoothing, beyond that produced by the telescope, is present.

The results of initial phase refinement at frequencies less than 25 cycles/arc-second, followed by super-resolution, are shown in Figs. 27 a-d), and in Fig. 28. The need for a final digital lowpass filter has been eliminated, since the algorithm forces the solution to be free of ringing. This is perhaps the primary effect of super-resolution on this reconstruction, since estimating the highest frequencies increased the total energy by only about 3%. The image shows little evidence of fine surface structure, in contrast to the result of McDonnell and Bates [7].

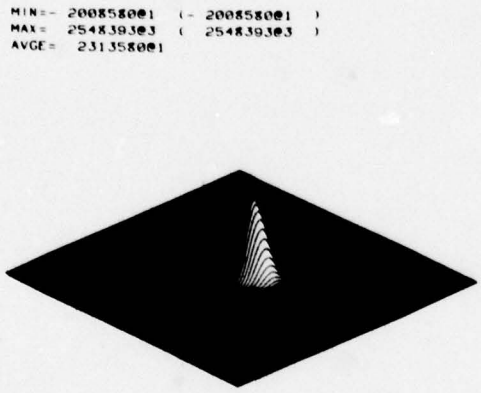
As discussed in chapter IV, there are some internal evidences that can be examined to evaluate the accuracy of the reconstruction. First, the sum squared difference between the measured spectrum and the super-resolved spectrum is about 2.5% of the total energy of those frequencies within the passband of the telescope. This is consistent with the results obtained in the simulations



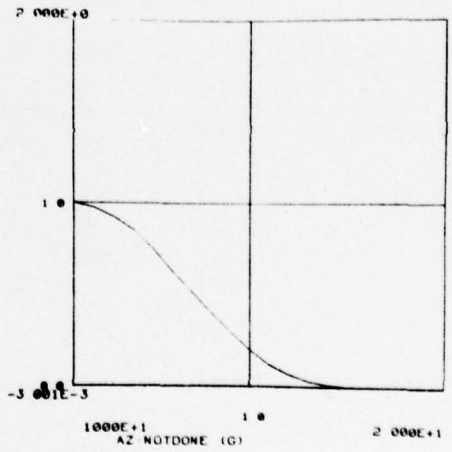
a) Magnitude Spectrum



b) Phase Spectrum



c) Intensity Object



d) Average Radial Intensity Profile

Fig. 27 Super-Resolved Betelgeuse Reconstruction

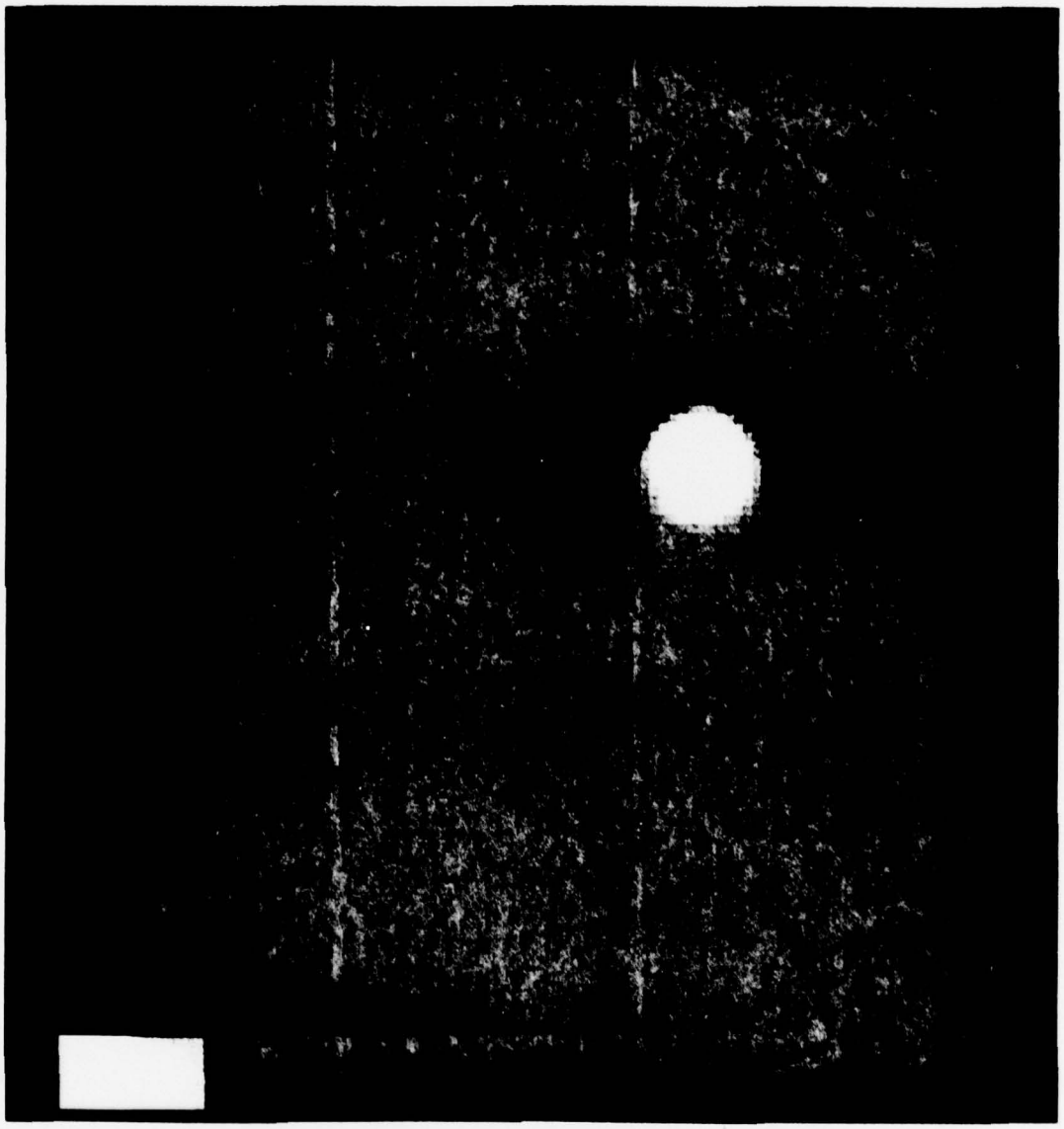


Fig. 28 Super-Resolved Betelgeuse Intensity Object

using noise-free data. Second, the super-resolved magnitude spectrum decays nicely with increasing frequency, and shows no suspect structure at the high frequencies. Third, the diameter of the reconstruction, which is .065 + .005 arc-seconds, agrees very well with the autocorrelation estimate derived from the unwindowed speckle images, which has a diameter of .116 arc-seconds. The reconstruction has been examined by astronomers, who indicate that the average radial intensity profile of Fig. 27 d) is in excellent agreement with astrophysical models of the star. Fourth, the peak negative light intensity in the interpolated image is only about 1% of the peak positive light intensity. Finally, because the spatial mask was square of width 7 pixels, it cannot be claimed that the spatial mask produced the symmetry of the resulting reconstruction. The fact that the reconstruction has circular symmetry when the spatial mask approximated so poorly the region of support is due undoubtedly to the fact that the spectrum was very nearly resolved by the telescope. Thus the primary effect of the processing was to provide an accurate phase estimate.

V.3 Capella reconstruction

The reconstruction of Capella was less direct than that of Betelgeuse because of the difficulty in determining the size and spacing of the spatial masks. The Labeyrie autocorrelation estimate is so severely filtered that only

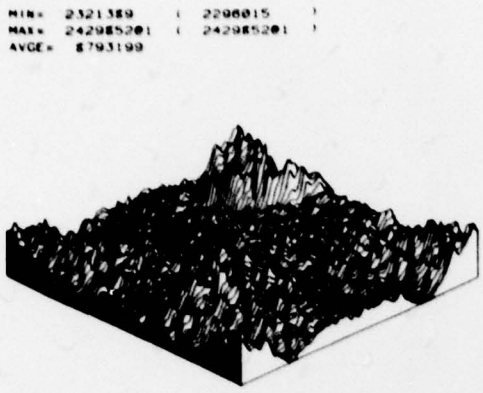
the relative orientation of the two stars and a very loose bound on the area of the region of support could be obtained from it. The sizes and spacing of the final mask were arrived at by trial and error and represent the best fit to the measured frequency spectrum. I used as a goodness-of-fit criteria the following parameters: 1) the difference in energy between the initial spectrum and the corresponding lowpass portion of the extrapolated spectrum, 2) the location and magnitude of the first spectral peak and valley, and 3) the negativity in the reconstructed image after it had been interpolated by a factor of 4. It was necessary in the processing to obtain finer resolution of the image in order to adequately approximate the size of the star discs. The measured frequency spectrum was low pass filtered with a separable Fourier window with cutoff frequency of 27 cycles/arc-second. The final spatial mask used two radially symmetric masks of diameter 5 pixels spaced diagonally from each other 8 pixels over and up. This corresponds to an interstar distance of .028 arc-seconds, and a common star diameter of about .0125 arc-seconds.

The spectra illustrated in Figs. 29 a-b) were obtained from 45 Hamming windowed speckle photographs. The profile of the magnitude spectrum taken along the antidiagonal makes it easier to spot the telescope cutoff frequency, at about 69 cycles/picture (27 cycles/arc-second), as shown in

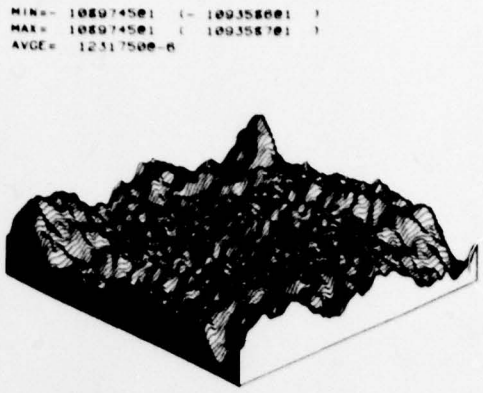
Fig. 29 c). Note that the phase estimate has failed except at the lowest frequencies. The Labeyrie autocorrelation estimate is shown in Fig. 29 d). The resolution is so poor that a meaningful image cannot be formed from this data. The super-resolved spectrum is shown in Figs. 30 a-c). The energy in the spectrum is now 4.18 times the originally measured energy, and the highest extrapolated frequencies are now at 147 cycles/arc-second (resolution equivalent to an 800-inch telescope). The first spectral valley and peak are in essentially the same locations, and the SSE between the measured spectrum and the corresponding portion of the extrapolated spectrum is 2.7%. The smoothing effects of the algorithm are especially evident in the antidiagonal profile of the extrapolated spectrum. The reconstructed object, interpolated by a factor of 16, is shown as a line-drawing in Fig. 30 d) and as an intensity object in Fig. 31. The object had a maximum negative value 2.85% of the peak positive intensity. The two stars have an intensity ratio of 1:.638. This value is given in [60] as 1:.632. The reconstruction thus appears to be reasonably accurate.

It is obvious in inspecting the reconstruction that the spatial masks have interfered with estimating the star discs. Although the dim star diameter is about correct, the bright star diameter has definitely been underestimated. In [61] it is reported that the bright and

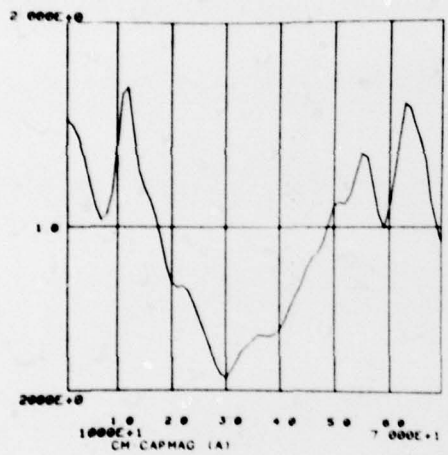
dim stars have diameters of 13 and 7 times the solar diameter. It thus appears that even this artifact is consistent with current astrophysical measurements. Because the spatial mask has in this case such a strong influence on the reconstruction, one would like to exhaustively explore the effects of variations, such as using masks of slightly different sizes for the two star discs. It is likely that better results can be obtained by such experiments. Unfortunately, really fine spatial resolution would require huge arrays, and is entirely beyond the practical capabilities of our computing facilities.



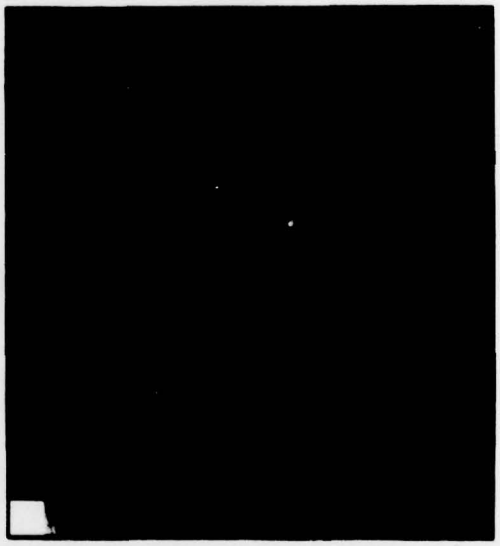
a) Labeyrie Magnitude Spectrum Estimate



b) Least-Squares Phase Estimate

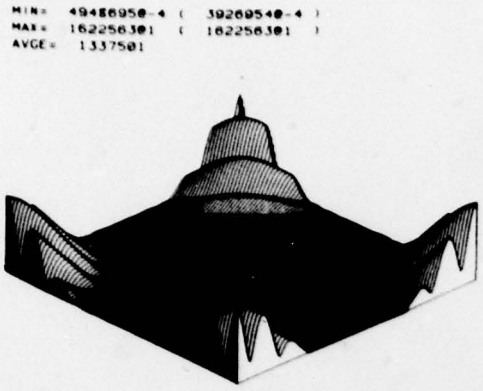


c) Anti-Diagonal Profile of Magnitude Spectrum to 70 cycles/picture

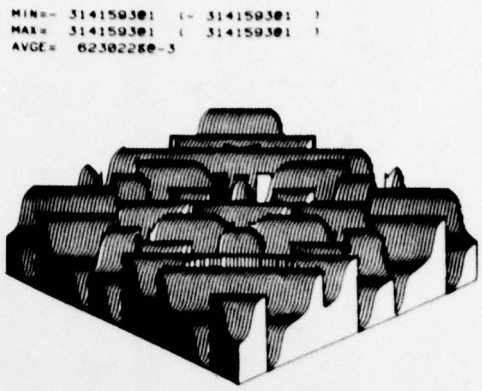


d) Labeyrie Auto-Correlation Estimate

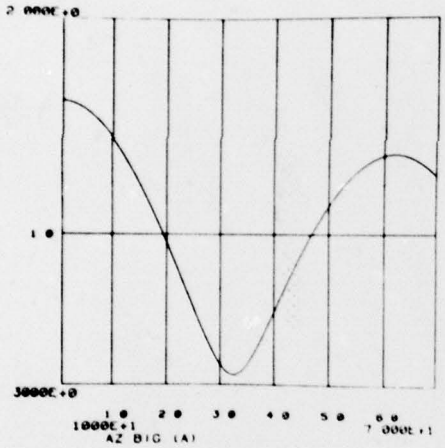
Fig. 29 Capella Reconstructions Obtained from Hamming Windowed Speckle Photographs



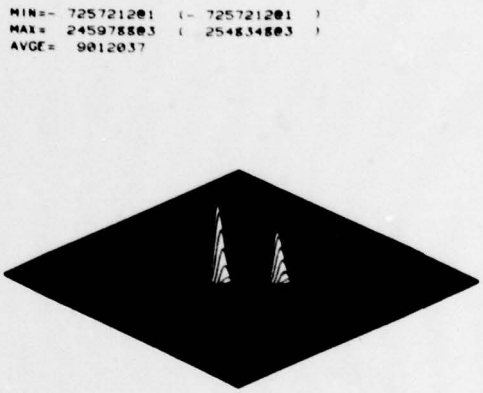
a) Magnitude Spectrum



b) Phase Spectrum



c) Anti-diagonal Profile of Magnitude Spectrum to 70 cycles/picture



d) Intensity Object

Fig. 30 Super-Resolved Capella Reconstruction

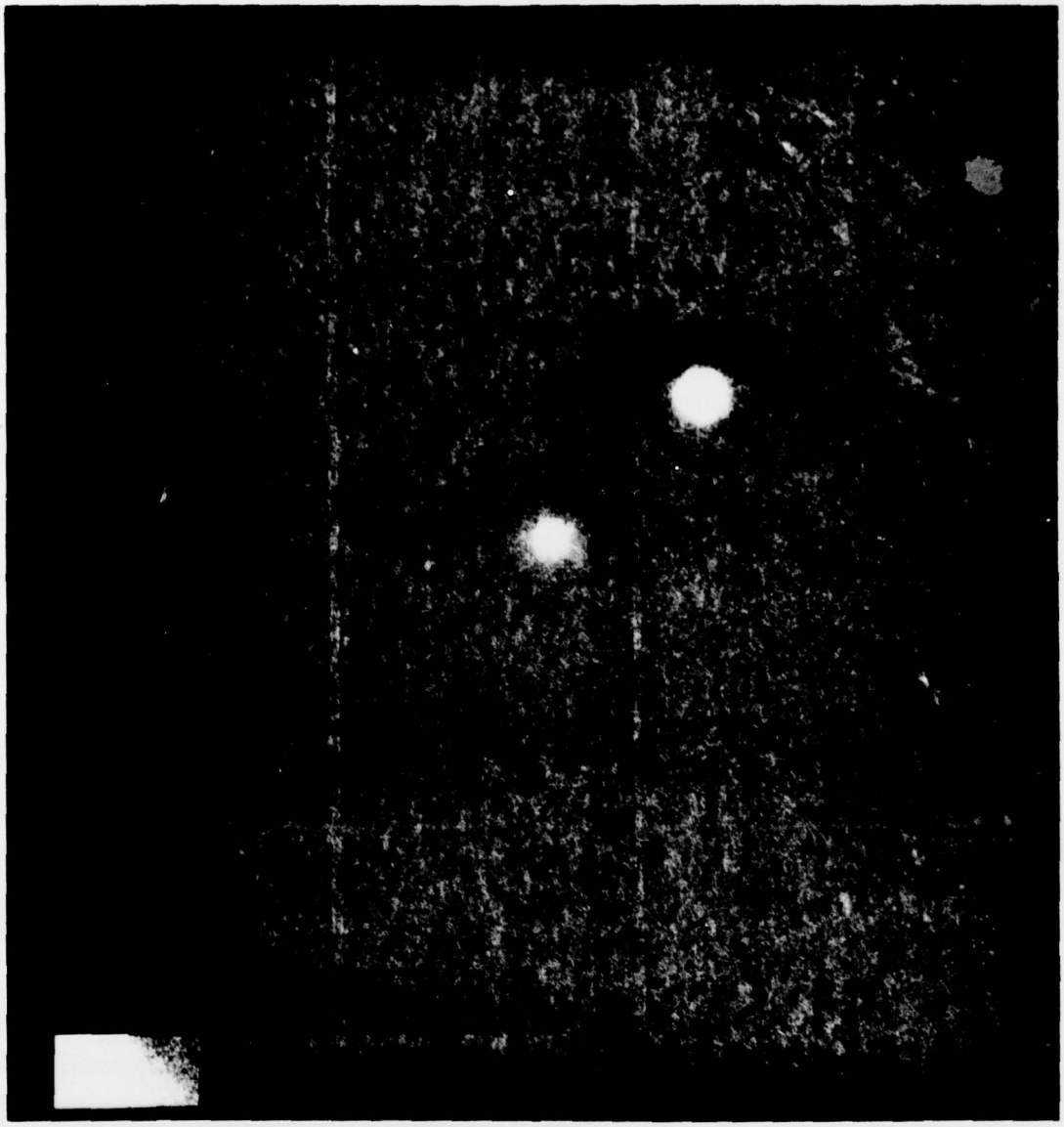


Fig. 31 Super-Resolved Capella Intensity Object

CHAPTER VI

SUMMARY AND CONCLUSIONS

As claimed in the introduction, we have presented a method of obtaining reliable phase spectrum estimates for objects blurred by atmospheric turbulence. The method is based on the Knox-Thompson phase difference estimator which, without modification, seems unreliable and overly sensitive to noise. We have also presented a new super-resolution algorithm which appears to have the best stability and convergence properties of any such algorithm proposed to date. Both methods have been applied to simulated and real blurred stars with substantial success.

As with most other new techniques, however, a number of issues, both pragmatic and theoretical, need further study. The first obvious question relates to the accuracy of the reconstructions of Betelgeuse and Capella. There is some recent evidence [52] which suggests that our speckle photographs are not of the highest quality. Specifically, one would expect better results if exposure times were about 10 ms, rather than 50 ms. This would necessitate the use of wider bandpass optical interference filters to permit adequate exposure of the speckle images, say 100 Å instead of the 20 Å of our data. Certainly, a study of the

effects of the various exposure time-filter bandwidth tradeoffs should be made for astronomical applications. Furthermore, solid-state cameras are now available which record a sampled and digitized image directly onto magnetic tape. This promises important reductions in sensor noise level in the data. Consequently, we feel that although our reconstructions are probably the best to date, they cannot be considered as the last word on the astrophysical structure of these stars. Rather, they indicate the power of these new processing techniques. A possible empirical check on these techniques would be to reconstruct an object for which we have a satellite photograph uncorrupted by atmospheric turbulence. The Galilean moons of Jupiter are possible candidates. If our results are confirmed by further research, an exciting astronomical application of the new super-resolution algorithm would be to data from the orbital telescope now under construction.

Another research area is the application of these methods to spectral estimation, bandwidth compression, and blind deconvolution problems for both one- and two-dimensional signals. In this latter problem, one requires a prototype spectrum so as to be able to estimate the spectrum of the blur, which is then removed from the blurred signal by some inverse filtering technique. We remind the reader of the importance of an 'adequate' initial phase spectrum estimate, which may inhibit the

application of the super-resolution technique to certain problems.

There is no doubt that our present theoretical understanding of these techniques is incomplete. For example, rather detailed studies of the bias and variance characteristics, and how these are modified by analysis windows, have been made for PSD estimators. For completeness, similar studies should be made for the Knox-Thompson estimator. However, I feel an even more important development would be the creation of a general analytical framework in which to cast the super-resolution algorithm (which includes the phase refinement algorithm). Youla's work [48] is probably the foundation of such a framework. Our algorithm appears to be a special case of the regularizors discussed in that paper. The analysis is complicated by the non-linearities of the algorithm, however. Hopefully, such an analysis would provide a quantitative description of the effects of noise and the rate of frequency extrapolation on the final estimates. In fact, such an analysis may indicate that the algorithm presented here is not optimal. This would indicate the possibility of even better results than those presented here. As a practical matter, however, the simplicity of the implementation and the rapidity of convergence of this new algorithm may make it a processing tool of importance.

REFERENCES

[1] A. Labeyrie, "Attainment of Diffraction-Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. & Astrophys., vol. 5, no. 1, May 1970, pp. 85-87.

[2] D.Y. Gezari, A. Labeyrie, and R.V. Stachnik, "Speckle Interferometry: Diffraction-Limited Measurements of Nine Stars with the 200 Inch Telescope," The Astrophys. Jour., vol. 173, no. 1, April 1972, pp. L1-L5.

[3] R.H.T. Bates, P.T. Gough, and P.J. Napier, "Speckle Interferometry gives Holograms of Multiple Star Systems," Astr. and Ap., vol. 22, no.1, 1973, pp. 319-320.

[4] C.Y.C. Liu and A.W. Lohmann, "High Resolution Image Formation through the Turbulent Atmosphere," Opt. Comm., vol. 8, no. 2, Aug. 1973, pp. 372-377.

[5] G.P. Weigelt, "Large Field Speckle Interferometry," Optik, vol. 43, no. 2, Aug. 1975, pp. 111-128.

[6] C.R. Lynds, S.P. Worden, and J.W. Harvey, "Digital Image Reconstruction Applied to Alpha Orionis," The Astrophys. Jour., vol. 207, no. 1, July 1976, pp. 174-180.

[7] M.J. McDonnell and R.H.T. Bates, "Digital Restoration of an Image of Betelgeuse," The Astrophys. Jour.,vol. 208, no. 2, Sept. 1976, pp. 443-452.

[8] G.L. Welter and S.P. Worden, "A Method for Processing Stellar Speckle Interferometry Data," J. Opt. Soc. Am., vol. 68, no. 9, Sept. 1978, pp. 1271-1275.

[9] J.W. Sherman, "A Posteriori Restoration of Atmospherically Degraded Images using Multiframe Imagery," Proc. SPIE, vol. 74, 1976, pp. 249-258.

[10] J.W. Sherman, "Speckle Imaging using the Principal Value Decomposition Method," to be published.

[11] E.R. Cole, "The Removal of Unknown Image Blurs by Homomorphic Filtering," ARPA technical report UTEC-CSc-74-029.

[12] T.M. Cannon, "Digital Image Deblurring by Non-linear Homomorphic Filtering," ARPA technical report UTEC-CSc-74-091.

[13] B.L. McGlamery, "Image Restoration Techniques Applied to Astronomical Photography," in "Astronomical Use of Television-type Image Sensors," report No. N71-28509-625 of the National Technical Information Service of the U.S. Dept. of Commerce, 1971.

[14] K.T. Knox and B.J. Thompson, "Recovery of Images from Atmospherically-Degraded Short-Exposure Photographs," Astrophys. J., vol. 193, no. 1, Oct. 1974, pp. L45-48.

[15] K. Knox, "Diffraction-Limited Imaging with Astronomical Telescopes," PhD. Dissertation, Institute of Optics, The University of Rochester, Rochester, New York. (unpublished), 1975.

[16] F. Roddier, "Speckle Interferometry through Small Multiple Aperture Telescopes: Michelson Stellar Interferometry and Aperture Synthesis in Optics," Opt. Comm., vol. 10, no. 1, Jan. 1974, pp.103-105.

[17] P. Nisenson, D.C. Ehn, and R.V. Stachnik, "Astronomical Speckle Imaging," Proc. SPIE, vol. 75, 1976, pp. 83-88.

[18] R.V. Stachnik et al., "Speckle Image Reconstruction of Solar Features," Nature, vol. 266, no. 5597, March 1977, pp. 149-151.

[19] M. Rimmer, "Method for evaluating a Lateral Shear Interferometer," Appl. Opt., vol. 13, no. 3, Mar. 1974, pp. 623-629.

[20] D.L. Fried, "Least-square fitting a Wave-front Distortion Estimate to an Array of Noisy Phase Differences," J. Opt. Soc. Am., vol. 67, no. 3, March 1977, pp. 370-375.

[21] R.H. Hudgin, "Wave-front Reconstruction for Compensated Imaging," J. Opt. Soc. Am., vol. 67, no. 3, March 1977, pp. 375-380.

[22] J.W. Hardy, J.E. Lefebvre, and C.L. Koliopoulos, "Real-time Atmospheric Compensation," J. Opt. Soc. Am., vol. 67, no. 3, March 1977, pp. 360-369.

[23] B.R. Hunt, "Matrix Formulation of the Reconstruction of Phase Values from Phase Differences," J. Opt. Soc. Am., to be published.

[24] B.L. Buzbee, G.H. Golub, and C.W. Nielson, "On Direct Methods for Solving Poisson's Equation," SIAM J. Numer. Anal., vol. 7, no. 4, Dec. 1970, pp. 627-656.

[25] F.B.A. Giwa, "An Unconditionally Stable Scheme for the Numerical Solution of Finite Difference Forms of Poisson's Equation," J. Inst. Math. Applics., vol. 21, no. 4, June 1978, pp. 479-491.

[26] R. W. Hockney, "A Fast Direct Solution of Poisson's Equation using Fourier Analysis," Jour. Assoc. Comp. Mach., vol. 12, no. 1, Jan. 1965, pp. 95-113.

[27] R. W. Hockney, "The Potential Calculation and some Applications," Methods in Computational Physics, vol. 9, Academic Press, New York, 1970, pp. 135-211.

[28] D. Fischer et al., "On Fourier-Toeplitz Methods for Separable Elliptic Problems," Math. of Compt., vol. 28, no. 126, April 1974, pp. 349-368.

[29] T. Muir, Theory of Determinants, vol. 3, Dover Publications, Inc., New York, 1960.

[30] R.H.T. Bates, "On Phase Problems.I," Optik, vol. 51, no. 2, 1978, pp. 161-170.

[31] S.R. Robinson, "On the Problem of Phase from Intensity Measurements," J. Opt. Soc. Am., vol. 68, no. 1, Jan. 1978, pp. 87-92.

[32] D. Kohler and L. Mandel, "Source Reconstruction from the Modulus of the Correlation Function: A Practical Approach to the Phase Problem of Optical Coherence Theory," J. Opt. Soc. Am., vol. 63, no. 2, Feb. 1973, pp. 126-134.

[33] E. Wolf, "Is a Complete Determination of the Energy Spectrum of Light Possible from Measurements of the Degree of Coherence?," Proc. Phys. Soc. Lond., vol. 80, no. 2, Dec. 1962, pp. 1269-1272.

[34] E.L. O'Neill and A. Walther, "The Question of Phase in Image Formation," Opt. Acta., vol. 10, no. 1, Jan. 1963, pp. 33-40.

[35] A. Walther, "The Question of Phase Retrieval in Optics," Opt. Acta., vol. 10, no. 1, Jan 1963, pp. 41-49.

[36] A.J. Devaney and R. Chidlaw, "On the Uniqueness Question in the Problem of Phase Retrieval from Intensity Measurements," J. Opt. Soc. Am., vol. 68, no. 10, Oct. 1978, pp. 1352-1354.

[37] R.W. Gerchberg and W.O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik, vol. 35, no. 2, April 1972, pp. 237-246.

[38] A.M.J. Huiser et al., "On Phase Retrieval in Electron Microscopy from Image and Diffraction Pattern," Optik, vol. 45, no. 4, July 1976, pp. 303-316.

[39] J. Gassmann, "Optimal Iterative Phase Retrieval from Image and Diffraction Intensities," Optik, vol. 48, no. 4, Aug. 1977, pp. 347-356.

[40] J.R. Feinup, "Reconstruction of an Object from the Modulus of its Fourier Transform," Opt. Lett., vol. 3, no. 1, July 1978, pp. 27-29.

[41] R.H.T. Bates, "On Phase Problems.II," Optik, vol. 51, no. 3, Jan. 1979, pp. 223-234.

[42] J.L. Harris, "Diffraction and Resolving Power," J. Opt. Soc. Am., vol. 54, no. 7, July 1964, pp. 931-936.

[43] C.W. Barnes, "Object Restoration in a Diffraction-Limited Imaging System," J. Opt. Soc. Am., vol. 56, no. 5, May 1966, pp. 575-578.

[44] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V. The Discrete Case," Bell Sys. Tech. Jour., vol. 57, no. 5, May-June 1978, pp. 1371-1430.

[45] C.K. Rushforth and R.W. Harris, "Restoration, Resolution, and Noise," J. Opt. Soc. Am., vol. 58, no. 4, April 1968, pp. 539-545.

[46] B.R. Frieden, "Restoring with Maximum Likelihood and Entropy," J. Opt. Soc. Am., vol. 62, no. 4,

April 1972, pp. 511-618.

[47] R.W. Gerchberg, "Super-resolution through Error-Energy Reduction," Optica Acta, vol. 21, no. 9, 1974, pp. 709-720.

[48] D.C. Youla, "Generalized Image Restoration by the Method of Alternating Orthogonal Projections," IEEE Trans. Circ. Sys., vol. 25, no. 9, Sept. 1978, pp. 694-702.

[49] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," IEEE Trans. Circ. Sys., vol. 22, no. 9, Sept. 1975, pp. 735-742.

[50] J.A. Cadzow, "An Extrapolation Procedure for Band-Limited Signals," IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-27, no. 1, Feb. 1979, pp. 4-12.

[51] G. Toraldo Di Francia, "Resolving Power and Information," J. Opt. Soc. Am., vol. 45, no. 7, July 1955, pp. 497-601.

[52] D.P. Karo and A.M. Schneiderman, "Speckle Interferometry at Finite Spectral Bandwidths and Exposure Times," J. Opt. Soc. Am., vol. 68, no. 4, April 1978, pp. 480-485.

[53] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, Inc., San Francisco, 1968.

[54] S.K. Mitra and M.P. Ekstrom, ed., Two-Dimensional Digital Signal Processing, Dowden, Hutchinson, & Ross, Inc., Stroudsburg, Pennsylvania, 1978.

[55] H. C. Andrews and B.R. Hunt, Digital Image Reconstruction, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977.

[56] G.M. Jenkins and D.G. Watts, Spectral Analysis and Its Applications, Holden-Day, Inc., San Francisco, 1968.

[57] L.H. Koopmans, Spectral Analysis of Time Series, Academic Press, New York, 1974.

[58] A.V. Oppenheim and R.W. Schaefer, Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[59] F.J. Flanagan, Complex Variables: Harmonic and Analytic Functions, Allyn and Bacon, Inc., Boston,

1972.

[60] R.G. Aitken, The Binary Stars, McGraw-Hill, Inc., New York, 1935.

[61] R. Burnham, Jr., Burnham's Celestial Handbook, vol. 1, Dover, Inc., New York, 1978.